Short-time multichannel noise power spectral density estimators for acoustic signals

By

Jonathan Blanchette

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.A.Sc. in Electrical and Computer Engineering

Ottawa-Carleton Institute for
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Résumé

L'estimation de la densité spectrale de puissance est une étape critique dans plusieurs algorithmes de rehaussement de la parole. La demande pour des systèmes de rehaussement à canaux multiples est grande pour de nombreuses applications par exemple pour des systèmes de téléconférence, de téléphonie cellulaire, et pour les appareils auditifs. Le premier objectif de cette thèse est de développer un cadre général à canaux multiples pour résoudre la densité spectrale de puissance du bruit diffus lorsque la matrice de corrélation ou de cohérence spatiale est pré-estimée et lorsque le nombre d'interlocuteurs est moins élevé que le nombre de microphones. Le second objectif est de poursuivre le développement des solutions de forme analytique. La performance des algorithmes développés est évaluée en comparant leur précision avec des algorithmes préexistants et en utilisant des mesures de performances prescrites.

Abstract

The estimation of power spectral densities is a critical step in many speech enhancement algorithms. The demand for multi-channel speech enhancement systems is high with applications in teleconferencing, cellular phones, and hearing aids. The first objective of the thesis is to develop a general multi-channel framework to solve for the diffuse noise power spectral densities whenever the spatial correlation or coherence matrix is pre-estimated and the number of speakers is less than the number of microphones. The second objective is to develop closed-form analytical solutions. The performance of the developed algorithms is evaluated with pre-existing algorithms using prescribed performance measures.

Key words:

Noise power spectra estimation, diffuse noise field, multichannel acoustic system, speech enhancement, subspace decomposition

Acknowledgements

I am very thankful to my supervisor Dr. Martin Bouchard for giving me the opportunity to work on the topic of the thesis. Throughout my graduate studies he offered his guidance and helpful comments.

I would like to thank my beautiful wife Judith for her love and support. I appreciate when she forced me into bed when I was working late. Because of her, I have not turned into a computer chair potato.

I am grateful for my family`s support, except for my great uncle who jokingly laughs at me because Judith is the only one in the couple that "really" works.

Finally, a special thanks to my brothers and friends for impeding me from becoming a hermit and for keeping me in shape (*sauf pour mon foie*).

Original/Novel ideas contained in the thesis

In the table of contents and in the section titles, the novelties are marked with a bracketed star (★).

Contents

List of figures

List of tables

List of acronyms

AIC ..... Akaike Information Criterion
BTE ..... Behind-the-ear
CHT ..... Cayley-Hamilton theorem
DTFT ..... Discrete time Fourier transform
FFT ..... Fast Fourier transform
HRIR ..... Head related impulse response
HRTF ..... Head related transfer function
i.i.d. ..... Independent identically distributed
ILD ..... Interaural level difference
iSNR ..... Input signal to noise ratio
ITD ..... Interaural time difference
MDL ..... Minimum description length
MIMO ..... Multiple-input multiple-output
ML ..... Maximum likelihood
MS ..... Minimum statistics
MVDR ..... Minimum variance distortionless response
PLDNE ..... Power level difference noise estimator
PSD ..... Power spectral density
SHT ..... Sequential hypothesis test
SNR ..... Signal to noise ratio
STFT ..... Short-time Fourier transform
VAD ..... Voice activity detection
WSS ..... Wide sense stationary

List of symbols

This list does not contain the symbols presented in the Minimum Statistics and the Minimum Mean Square Error algorithms of section 3.1, and the Binaural approach algorithm of section 3.2.

Latin alphabet

$\boldsymbol{a}, \boldsymbol{a}(\theta, \phi)$	Propagation vector (p.6)
A	Total area of surface irradiating the noise (p.19)
$AIC(\widehat{N})$	Akaike Information Criterion (p.31)
$\boldsymbol{B}^A$	Adjoint of matrix $\boldsymbol{B}$ (p.33)
$b_n(x)$	Term associated to the n^{th} harmonic of the total pressure spectrum (p.16)
c	Speed of sound (p.6)
$C_n(\omega)$	Term associated to the n^{th} harmonic of the outgoing pressure spectrum (p.15)
$cord(a, b)$	Chordal metric of a with b (p.29)
$d(\hat{\boldsymbol{\Gamma}}_\eta, \boldsymbol{\Gamma}_\eta)$	Multichannel log norm of the error of the PSD estimate (p.26)
$d_{dB}(\hat{\boldsymbol{\Gamma}}_\eta, \boldsymbol{\Gamma}_\eta)$	Multichannel log norm of the error of the PSD estimate in dB (p.26)
$\mathbf{D}_{\text{mod}}$	Modified diagonal matrix of noise auto-PSDs (p.56)
$\boldsymbol{D}(\omega)$	Diagonal matrix built with diagonal terms of $\boldsymbol{\Gamma}_\eta(\omega)$ (p.11)
$def(\boldsymbol{\Gamma}_x)$	Defect number of matrix $\boldsymbol{A}_{M\times M}$, i.e., $M - rank(\boldsymbol{A})$, or the dimension of the null space of $\boldsymbol{A}$ that is, $\dim(\mathcal{N}(\boldsymbol{A}))$ (p.27)
$\boldsymbol{e}_n$	Vector with unit n^{th} entry, all the other entries are null (p.24)
$\mathcal{E}(Z)$	Expected value operation on random variable Z (p.7)
$\mathcal{E}(Z\|W)$	Conditional expectation of Z given W (p.21)
$\boldsymbol{G}$	Square root Wiener filter matrix (p.55)
$h_i(t, \theta, \phi)$	Transfer function of the point at angles (θ, ϕ) of the surface to the i^{th} microphone (p.10)
$h_{i,j}[n]$	Impulse response function from j^{th} source to i^{th} microphone (p.9)
$h_n^{(1)}(z), h_n(z)$	Hankel function of the first kind (p.15)
$\boldsymbol{h}[n]$	Impulse response vector (p.7)
$\boldsymbol{h}_{M\times N}[n], \boldsymbol{h}[n]$	Impulse response matrix (p.9)
$H_l(\omega)$	Left HRTF (p.8)
$H_r(\omega)$	Right HRTF (p.8)
$H_i(\omega, \theta)$	Transfer function angle θ to i^{th} microphone (p.20)
$H_i(\omega, \boldsymbol{\Omega})$	Transfer function from point on special coordinates $\boldsymbol{\Omega}$ to i^{th} microphone (p.19)
$H_{i,j}[r, k]$	Discrete STFT of impulse response of j^{th} source to i^{th} microphone (p.22)
$H_{i,j}(\omega)$	Transfer function of j^{th} source to i^{th} microphone (p.9)
$H_{W_{i,j}}$	Wiener filter to approximate the i^{th} noise spectrum from the j^{th} noise spectrum (p.21)
$\boldsymbol{H}_{M\times N}[r, k], \boldsymbol{H}_{M\times N}, \boldsymbol{H}$	Noisy signal spectrum vector in the discrete STFT domain (p.22)
$\boldsymbol{H}(\omega)$	Transfer function vector (p.8)
$\boldsymbol{H}_{M\times N}(\omega)$	Transfer function vector (p.9)
$\boldsymbol{H}(\omega, \theta)$	Transfer function vector from angle θ (p.20)
$\boldsymbol{H}(\omega, \theta, \phi)$	Transfer function vector (p.7)
$\boldsymbol{H}(\omega, \boldsymbol{\Omega})$	Transfer function vector from point on special coordinates $\boldsymbol{\Omega}$ (p.19)
$\boldsymbol{v}^H$	Hermitian transpose of $\boldsymbol{v}$ (p.7)
$iSNR_{dB}$	Input SNR in dB (p.24)
$\boldsymbol{I}$	Identity matrix (p.11)
J	Imaginary unit $\sqrt{-1}$ (p.4)
$j_n(z)$	Spherical Bessel function of order n (p.14)

$J_0(z)$	Bessel function of order 0 (p.19)
$\boldsymbol{k}$	Wavenumber vector (p.13)
$\mathcal{K}$	Used number of frequency bins in the cost function averaging (p.24)
L	Number of samples used in the noisy signal PSD matrix estimate (p.30)
L_w	Analysis window length (p.4)
$L_N(\widehat{N})$	Model order estimation sufficient statistic (p.31)
M	Number of microphones (p.6)
$MDL(\widehat{N})$	Minimum description length (p.32)
$MSE(\hat{\boldsymbol{\Gamma}}_\eta, \boldsymbol{\Gamma}_\eta)$	Mean Frobenius norm of the error of the PSD estimate (p.25)
$MSlogE(\hat{\boldsymbol{\Gamma}}_\eta, \boldsymbol{\Gamma}_\eta)$	Multichannel Frobenius norm of the error of the log PSD estimate (p.26)
N	Number of sources (p.9)
$N(\omega, \theta, \phi, \theta_i, \phi_i)$	Noise pressure spectrum at i^{th} microphone caused by point on surface located at angles (θ, ϕ) (p.13)
$N_i(\omega)$	Noise spectrum of i^{th} microphone (p.9)
$N_i[r, k]$	Discrete STFT of diffuse noise at i^{th} microphone (p.22)
N_{fft}	FFT length (p.4)
$\widehat{N}_{AIC}$	Estimated number of sources with the AIC (p.32)
$\widehat{N}_{SHT}$	Estimated number of sources with the sequential hypothesis testing (p.31)
$\boldsymbol{N}(\omega)$	Noise spectrum vector (p.9)
$\boldsymbol{N}[r, k], \boldsymbol{N}$	Diffuse noise spectrum vector in the discrete STFT domain (p.22)
$\mathcal{N}(\boldsymbol{A})$	Null space of $\boldsymbol{A}$ (p.27)
$O(x)$	Magnitude order of x (p.31)
$p_{in}(\omega, r, \theta, \phi, \theta_i, \phi_i)$	Incoming pressure spectrum at i^{th} microphone caused by point on surface located at angles (θ, ϕ) (p.13)
$p_{in}(\omega, r, \Theta_i)$	Incoming pressure spectrum at i^{th} microphone from point on spherical surface (p.15)
$p_{out}(\omega, r, \theta, \phi, \theta_i, \phi_i)$	Outgoing pressure spectrum at i^{th} microphone caused by point on surface located at angles (θ, ϕ) (p.13)
$p_{out}(\omega, r, \Theta_i)$	Outgoing pressure spectrum at i^{th} microphone from point on spherical surface (p.15)
$P_n^m(z)$	Associated Legendre function (p.14)
$\boldsymbol{P}_x, \boldsymbol{P}_{\eta^\perp}$	Signal subspace projection matrix (p.52)
$\boldsymbol{P}_\eta, \boldsymbol{P}_{x^\perp}$	Noise subspace projection matrix (p.52)
$rank(\boldsymbol{A})$	Rank of matrix $\boldsymbol{A}$ (p.27)
$\boldsymbol{r}_i$	Position of i^{th} microphone in Cartesian coordinates (p.6)
R	Window spacing (p.4)
$\boldsymbol{R}_{xx}(\tau)$	Correlation matrix of $\boldsymbol{x}(t)$ (p.7)
$\mathcal{R}$	Used number of time frames in the cost function averaging (p.24)
$\mathcal{R}(\boldsymbol{A})$	Space spanned by columns of $\boldsymbol{A}$ (p.29)
$\mathcal{R}^\perp(\boldsymbol{A})$	Space orthogonal to the range of $\boldsymbol{A}$ (p.29)
$s(t)$	Far field acoustic source signal (p.6)
$s[n]$	Source signal (p.7)
$s_j[n]$	Source signal measured by j^{th} microphone (p.9)
$\boldsymbol{s}[n]$	Sources vector (p.9)
$S(\omega)$	Spectrum of signal at source (p.7)
S^2	Surface domain (p.19)
$S_j(\omega)$	Spectrum of j^{th} source (p.9)
$S_j[r, k]$	Discrete STFT of j^{th} source (p.22)

$\boldsymbol{S}(\omega)$	Sources spectrum vector (p.9)
$tr(\boldsymbol{A})$	Trace of matrix $\boldsymbol{A}$ (p.25)
$\boldsymbol{v}^T$	Transpose of $\boldsymbol{v}$ (p.6)
$V(\omega)$	Spectrum of noise emanating from surface (p.13)
$\boldsymbol{V}$	Eigenvector matrix (p.50)
$\boldsymbol{V}_x$	Eigenvector matrix that span the signal subspace (p.50)
$\boldsymbol{V}_\eta$	Eigenvector matrix that span the noise subspace (p.50)
$\boldsymbol{V}(\omega)$	Unitary matrix in which the columns are the right eigenvalues of $\boldsymbol{\Gamma}_\eta(\omega)$ (p.21)
$w[m]$	Analysis window (p.4)
$\boldsymbol{w}_n$	n^{th} column of $\boldsymbol{W}_{M\times M}$ (p.24)
$\boldsymbol{W}_{M\times M}$	Wiener filter matrix to approximate the sources at microphones (p.23)
$\boldsymbol{\mathcal{W}}(\omega)$	WSS frequency dependent white Gaussian noise (p.22)
$x[r,m]$	Windowed sequence of $x[n]$ (p.4)
$x_i[n]$	Sources signals received at i^{th} microphone (p.9)
$x_i(t)$	Source measured at i^{th} microphone (p.6)
$\boldsymbol{x}[n]$	Sources received at microphones vector (p.9)
$x_l[n]$	Source signal heard at left ear (p.7)
$x_r[n]$	Source signal heard at right ear (p.7)
$X[n,k]$	Discrete STFT of $x[n]$l (p.4)
$X_l(\omega)$	Signal spectrum at left ear (p.8)
$X_i(\omega)$	Spectrum of signal at i^{th} microphone (p.7)
$X_i[r,k]$	Discrete STFT of sources at i^{th} microphone (p.22)
$X_r(\omega)$	Signal spectrum at right ear (p.8)
$X_{stft}[n,\omega)$	Short time Fourier transform of sampled signal $x[n]$ (p.4)
$\boldsymbol{X}(\omega)$	Spectrum of signal vector (p.7)
$\boldsymbol{X}[r,k],\boldsymbol{X}$	Sources sampled at microphones spectrum vector in the discrete STFT domain (p.22)
$\hat{x}[n]$	Reconstructed signal with overlap-add method (p.4)
$\hat{X}_n$	Estimate of the n^{th} entry of $\boldsymbol{X}$ (p.24)
$\widehat{\boldsymbol{X}}$	Estimate of $\boldsymbol{X}$ (p.23)
$\lvert X[r,k]\rvert^2$	Spectrogram of $x[n]$ (p.5)
$y_i[n]$	Noisy signal at i^{th} microphone (p.9)
$\boldsymbol{y}[n]$	Noisy signal vector (p.9)
$Y_i[r,k]$	Discrete STFT of noisy signal at i^{th} microphone (p. 22)
$Y_i(\omega)$	Noisy signal spectrum of i^{th} microphone (p.9)
Y_n^m	Spherical harmonic of order n (p.14)
$\boldsymbol{Y}[r,k],\boldsymbol{Y}$	Noisy signal spectrum vector in the discrete STFT domain (p.22)
$\boldsymbol{Y}(\omega)$	Noisy signals spectrum vector (p.9)
Greek alphabet	
α	Recursive averaging constant (p.5)
$\alpha_i(\theta,\phi)$	Attenuation of source signal at the i^{th} microphone (p.6)
$\alpha_1,\alpha_2,\alpha_3$	Smoothing factors (p.49)
β,γ	Parameters for the Doblinger algorithm (p.41)
$\gamma_{ss}(t)$	Autocorrelation of $s(t)$ (p.7)
$\gamma_s[n]$	Autocorrelation function of $s[n]$ (p.8)
$\gamma_{99}^{(N)}$	Threshold used in the sequential hypothesis testing with a 99% interval of confidence (p.31)
$\Gamma_{ss}(\omega),\Gamma_s(\omega)$	PSD of $s(t)$ (p.7)
$\Gamma_{\eta_i}(\omega),\Gamma_{\eta_i\eta_i}(\omega)$	Diagonal terms of matrix $\boldsymbol{\Gamma}_\eta(\omega)$ with index i (p.11)

$\Gamma_{\eta_i\eta_j}(\omega)$	Cross-PSD terms in matrix $\mathbf{\Gamma}_{\eta}(\omega)$ with index (i,j) (p.11)
$\Gamma_{v}(\omega)$	Noise PSD of the noise emitted from surface (p.10)
$\Gamma_{\eta}^{(i)}$	i^{th} solution of $\chi_{\mathbf{\Gamma}_y,\mathbf{\Phi}}(\Gamma_{\eta}) = 0$, with solution sorted in descending order (p.32)
$\hat{\Gamma}_x[r,k]$	PSD estimate of $x[n]$ in the Discrete STFT domain (p.5)
$\mathbf{\Gamma}_s(\omega)$	Source signals PSD matrix (p.10)
$\boldsymbol{\Gamma}_s$	Sources matrix PSD in the discrete STFT domain (p.23)
$\boldsymbol{\Gamma}_x$	Sources sampled at microphones matrix PSD in the discrete STFT domain (p.23)
$\boldsymbol{\Gamma}_y$	Noisy signal matrix PSD in the discrete STFT domain (p.23)
$\boldsymbol{\Gamma}_{\eta}$	Diffuse noise matrix PSD in the discrete STFT domain (p.23)
$\mathbf{\Gamma}_x(\omega)$	PSD matrix of $x[n]$ (p.9)
$\mathbf{\Gamma}_{xx}(\omega,\theta,\phi)$	PSD matrix of $\boldsymbol{x}(t)$ (p.7)
$\mathbf{\Gamma}_y(\omega)$	Noisy signals PSD matrix (p.10)
$\hat{\mathbf{\Gamma}}_y[r,k]$	Noisy signal PSD matrix estimate (p.30)
$\mathbf{\Gamma}_{\eta}(\omega)$	Diffuse noise PSD matrix (p.10)
$\mathbf{\Gamma}_{\eta}(\omega,\theta,\phi)$	Noise PSD matrix of noise coming from point on surface at angles (θ,ϕ) (p.10)
$\hat{\mathbf{\Gamma}}_{\eta}^{(SC)}$	Single channel methods estimate of noise PSD matrix (p.60)
$\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$	First estimate of the noise PSD matrix (p.29)
$\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$	Second estimate of the noise PSD matrix (p.52)
$\widehat{\boldsymbol{\Gamma}}_{\eta}^{(3)}$	Third estimate of the noise PSD matrix (p.53)
δ	Small constant (p.61)
$\delta(n)$	Kroenecker delta function (p.16)
$\Delta\boldsymbol{v}_{ij}$	$\boldsymbol{v}_i - \boldsymbol{v}_j$ (p.7)
Δz_{ij}	$z_i - z_j$ (p.7)
$\Delta\Gamma_{PLDNE}[r,k]$	Power level difference noise estimator (p.48)
$\eta_i(t,\theta,\phi)$	Noise measured at i^{th} microphone coming from angles (θ,ϕ) of the surface (p.10)
$\eta_i[n]$	Diffuse noise at i^{th} microphone (p.9)
$\boldsymbol{\eta}[n]$	Diffuse noise vector (p.9)
θ	Elevation angle (p.6)
Θ_i	$\cos(\theta)\cos(\theta_i) + \sin(\theta)\sin(\theta_i)\cos(\phi - \phi_i)$ (p.15)
Θ_{ij}	$\cos(\theta_j)\cos(\theta_i) + \sin(\theta_j)\sin(\theta_i)\cos(\phi_j - \phi_i)$ (p.16)
λ_i	i^{th} eigenvalue in descending order (p.27)
$\lambda_i(\boldsymbol{A})$	i^{th} eigenvalue of matrix $\boldsymbol{A}$ sorted in descending order (p.28)
$\lambda_{min}(\boldsymbol{A},\mathbf{B})$	Minimal generalized eigenvalue, or minimal solution of $\chi_{\boldsymbol{A},\boldsymbol{B}}(\lambda) = 0$ (p.29)
$\mathbf{\Lambda}$	Eigenvalue matrix (p.50)
$\mathbf{\Lambda}(\omega)$	Positive definite diagonal eigenvalue matrix of $\mathbf{\Gamma}_{\eta}(\omega)$ (p.21)
$\mathbf{\Lambda}_{\mathrm{x}}$	Eigenvalues associated to the signal subspace (p.50)
$\mathbf{\Lambda}_{\eta}$	Eigenvalues associated to the signal subspace (p.50)
$v(t)$	Noise emanating from surface (p.10)
τ_i	Delay of source signal at the i^{th} microphone (p.6)
ϕ	Azimuth angle (p.6)
ϕ_{min}, ϕ_{max}	PLDNE thresholds (p. 49)
$\Phi_{\eta_i\eta_j}(\omega)$	Coherence functions between i^{th} and j^{th} channel (p.11)
$\mathbf{\Phi}(\omega)$	Coherence matrix (p.11)

$\boldsymbol{\Phi}_{\mathrm{mod}}$	Modified coherence matrix (p.56)
$\widehat{\boldsymbol{\Phi}}[r,k], \widehat{\boldsymbol{\Phi}}_r$	Coherence matrix estimate at time frame r (p.53)
$\widehat{\boldsymbol{\Phi}}'[r,k]$	Instantaneous coherence matrix estimate (p.53)
$\chi_{\boldsymbol{A},\boldsymbol{B}}(\lambda)$	Characteristic function of matrix pencil $\lambda\boldsymbol{B}-\boldsymbol{A}$,i.e., $\det(\lambda\boldsymbol{B}-\boldsymbol{A})$ (p.27)
$\boldsymbol{\Psi}(\omega), \boldsymbol{\Psi}$	Normalised spatial spectral matrix, correlation matrix (p.11)
$\widehat{\boldsymbol{\Psi}}[r,k], \widehat{\boldsymbol{\Psi}}_r$	Normalised spatial spectral matrix estimate at time frame r (p.57)
$\widehat{\boldsymbol{\Psi}}'[r,k]$	Instantaneous normalised spatial spectral matrix estimate (p.57)
$\boldsymbol{\Omega}$	Surface parameters (p.19)
Miscellaneous	
$\mathbf{1}$	Ones vector (p.46)
$\sqrt{\boldsymbol{A}}, \boldsymbol{A}^{1/2}$	Principal square root of matrix $\boldsymbol{A}$ (p.11)
$\boldsymbol{A} \circ \boldsymbol{B}$	Schur product of matrices $\boldsymbol{A}$ with $\boldsymbol{B}$ (p.11)
$f'(z)$	Derivative of $f(z)$ (p.15)
$s[n] * h[n]$	Linear convolution between $s[n]$ and $h[n]$ (p.7)
$\nabla_{\boldsymbol{v}^*}$	Conjugate gradient of vector $\boldsymbol{v}$ (p.29)
$\lvert z\rvert^2, \lvert\boldsymbol{v}\rvert^2$	Magnitude squared value of z or vector $\boldsymbol{v}$ (p.5)
z^*	Conjugate of z (p.14)

Chapter 1 Introduction

1.1 Motivation and objectives

The use of multiple microphones in speech processing is becoming more widespread. Many single and multi-channel beamforming or speech enhancement algorithms require to know the noise power spectral density (PSD) for example, the Wiener filter that estimates the sources from the microphones inputs (section 2.4.1). Developments in multichannel noise PSD estimation techniques have been made recently (e.g. [1-6]). There is still a lack of noise estimation algorithms for scenarios where there are multiple sources under non-stationary acoustic environments.

In general, diffuse noise fields can be inhomogeneous or homogeneous. A discussion on the nature of theses noise fields is presented in section 2.3. The main goal of this thesis is to develop a general framework to solve for the diffuse noise PSD matrix under the following problems/conditions:

- Possible presence of multiple sources under non-stationary acoustic environments.
    - The number of acoustic sources present is unknown.
- The transfer functions from the sources to the microphones are completely unknown.
    - No analytic function is used to model the transfer functions.
- Generalize previous work [3, 5] as subcases of a more general framework.
    - Find analytical solutions for the noise auto-PSD if they exist.
- The noise PSD matrix must be computed as precise as possible to maximize the performance of applications such as multichannel speech/sources enhancement.
- The noise auto-PSDs at the different microphones are almost surely different in practice.
    - Short-term differences occur even if the diffuse noise field is homogeneous.
    - Address the problem of using the homogeneous noise field framework instead of the correct inhomogeneous noise field model.
    - Profit from the estimated number of sources present in the environment to find the best noise PSD matrix when the auto-PSDs are not necessarily equal.
- Address the issue that if the environment changes, the models used to estimate the noise PSD matrix might be rendered obsolete.
- Exploit the information of highly correlated channels.

All the above issues are resolved to various degrees in the thesis, leading to various new contributions.

1.2 Thesis outline

The remainder of the thesis is organised as follows:

- Chapter 2 presents all the prior techniques used to develop the new algorithms. This includes:
    - A short discussion on the technique used to analyse and synthetize sampled signals, and the technique used to estimate the noisy signal power spectral density in section 2.1.
    - In section 2.2, starting with examples on the model used to describe sources signals in free field and with the head related perspective, we present the general system of equations that models multiple sources received at microphones under the presence of diffuse noise.

  - In section 2.3 we present a statistical and environment geometry dependent model for the noise PSD matrix. Definitions for the noise correlation and spatial spectral matrix are presented and derived for different scenarios such as cylindrical and spherical diffuse noise fields. A practical technique to compute the correlation and coherence matrices is provided in the end of the section.
  - In section 2.4, the basic equations concerning additive noise are reviewed with the discrete short time Fourier transform and the MIMO perspective. We put an emphasis on the importance of the noise PSD matrix estimation with a simple example of Wiener filtering. Various single and multichannel performance measures are presented and discussed in section 2.4.2.
  - In section 2.5, the algebraic solution for the noise PSD is presented when the number of sources is less than the number of microphones and the coherence matrix is available.
  - Section 2.6 contains a brief discussion on numerical sensitivity issues of the noise PSD estimates.
  - Section 2.7 presents the typical noisy signal algorithms to compute the sample noisy signal PSD matrix.
  - In section 2.8, we point out that if sample PSD matrices are used then it is absolutely necessary to estimate the approximate number of sources present, and use this information to get an unbiased estimate of the noise power level.
  - Finally in section 2.9, the analytic formulas for the ordered generalized eigenvalues are given. The noise power level is a combination of these eigenvalues and the number of eigenvalues used depends on the estimated number of signals.
- Chapter 3 presents various previously published single channel and multichannel noise PSD estimation algorithms.
- Chapter 4 presents the proposed algorithms used to estimate the noise PSD matrix.
  - Section 4.1, shows the full derivation of the algorithms that relate to the homogeneous noise field model (known coherence matrix).
  - By analogy, section 4.2 presents the same algorithms as in section 4.1 but when the correlation matrix is known instead of the coherence matrix.
  - In section 4.3, we discuss the fact that an algebraic solution of the noise power level is possible in the binaural scenario with a single source with known transfer functions. Although the solution exists, it cannot be used in practice since it does not apply to sample PSD matrices and it is too sensitive to model perturbations. This gives importance to the algorithm presented in section 4.1 as it solves the same problem of finding the various auto-PSDs under the condition that we have the coherence matrix to solve for an inhomogeneous noise field (noise field model mismatch).
  - In 4.4, a brief discussion on the minimum variance distortionless response solution for the noise power level is provided with the two noise field models (inhomogeneous and homogeneous).
- Chapter 5 presents simulation results comparing the proposed noise PSD estimation methods with single channel PSD estimation algorithms. Various parameters are tested to see their effect on the estimation precision such as the number of sources, the overall input SNR, and the models used for the noise field. Only single channel methods are used in the comparison because other existing multichannel algorithms suppose prior information on the sources transfer functions. This restriction is done to keep the comparison fair. Other algorithms that do not suppose information on the sources transfer

functions e.g., [3] can be included in our framework. This effectively excludes all multichannel algorithm presented in Chapter 3.

- Finally the conclusion, the contributions and the future works are discussed in Chapter 6.

It is worth mentioning that when this thesis assumes a binaural system where the information between the left and right sides is exchanged via a wireless links, a perfect transmission of all signals is assumed, i.e., for such systems the thesis does not take into consideration practical issues such as transmission delay of the microphone signals, available bandwidth and jitter, for example.

Chapter 2 Fundamentals

2.1 The short time Fourier transform

The short time Fourier transform (STFT) is given by [7]:

$$X_{stft}[n,\omega) = \sum_{m=0}^{L_w-1} x[m+n]w[m]e^{-\jmath m\omega}, \quad \textbf{2.1}$$

where $x[n]$ is the input signal, and $w[m]$ is an analysis window, ω is the radial frequency, and the dotless j "ȷ"is the imaginary unit $\sqrt{-1}$.

If we sample in time $X_{stft}[n,\omega)$ at every R samples and at N_{fft} equally spaced frequencies $\omega_k = 2\pi k/N_{fft}$, we can obtain a discrete version of the STFT. The window in Eq.(2.1) is set to a finite length L_w, and $L_w \leq N_{fft}$. The time-sampled discrete STFT $X[n,k]$ is defined to be [8]

$$X[r,k] = X_{stft}[rR,\omega_k) = \sum_{m=0}^{L_w-1} x[m+rR]w[m]e^{-\frac{\jmath 2\pi km}{N_{fft}}}, \quad \textbf{2.2}$$

where the frame $r \in]-\infty,\infty[$ and the frequency bin $k \in [0, N_{fft}-1]$. Choosing $R \leq L_w$ ensures that all samples of $x[n]$ are used. The term inside the DTFT of Eq.(2.2) is a windowed signal sequence

$$x[r,m] = x[m+rR]w[m], \quad \textbf{2.3}$$

where the time frame $r \in]-\infty,\infty[$, and $m \in [0, L_w-1]$. If ever we want to efficiently synthetize $x[n]$ from $X[r,k]$, it is possible to do so using the overlap-add method which consists of adding up reconstructed overlapping segments in such a way that they add up to the original signal $x[n]$. We first synthetise the segments $x[r,m]$ from $X[r,k]$ with the inverse fast Fourier transform (FFT) like the following

$$x[r,m] = \frac{1}{N_{fft}} \sum_{k=0}^{N_{fft}-1} X[r,k]e^{\frac{\jmath 2\pi km}{N_{fft}}}. \quad \textbf{2.4}$$

We subsequently add up the sequences to form a reconstructed signal $\hat{x}[n]$; $i.e.$,

$$\hat{x}[n] = \sum_{r=-\infty}^{\infty} x[r, n-rR]. \quad \textbf{2.5}$$

Inserting Eq.(2.3) into Eq.(2.5) we obtain the following:

$$\hat{x}[n] = \sum_{k=-\infty}^{\infty} x[n + rR - rR]w[n - rR] \\ = x[n] \sum_{k=-\infty}^{\infty} w[n - rR] = x[n]\tilde{w}[n]. \quad \textbf{2.6}$$

If $\tilde{w}[n] = C$ is a constant, then it is possible to perfectly recover $x[n]$ with:

$$x[n] = \hat{x}[n]/C \quad \textbf{2.7}$$

For example, perfect reconstruction is possible for the Bartlett window of length $L_w = M + 1$, M even, $R = M/2$[7, 8]. For the Bartlett window case, $C = 1$. Zero-padding the segments during the analysis phase allows linear convolution to be possible with filter $h[n]$ of length L_h. Because we implement the Fourier transform with the FFT, it is preferable to add up zeros to the segment in such a way that the FFT length becomes $N_{fft} = 2^{\nu+1}$, where $\nu = \lceil \log_2 L_w \rceil$ ("ceiling" function). This allows a filter length of $L_h \leq 2^{\nu+1} - L_w + 1$ without circular convolution or time aliasing effects.

2.1.1 PSD estimation from the STFT

An instantaneous power spectral density estimate or the spectrogram is defined as the squared magnitude of the discrete STFT:

$$|X[r,k]|^2 = 1/U \left| \sum_{m=0}^{L_w-1} x[m + rR]w[m]e^{-\frac{j2\pi km}{N_{fft}}} \right|^2. \quad \textbf{2.8}$$

with *U* being a normalization factor that accounts for the energy of the window function being used and its length. The use of *U* is required when it is important to preserve the physical units of the PSD. Throughout this thesis we simply use *U*=1. The power spectral density (PSD) estimate is then given by an average of successive spectrograms:

$$\hat{\Gamma}_x[r,k] = \frac{1}{L} \sum_{r=0}^{L-1} |X[r,k]|^2. \quad \textbf{2.9}$$

In practice [8] the PSD is computed recursively with

$$\hat{\Gamma}_x[r,k] = \alpha \Gamma_x[r-1,k] + (1-\alpha)|X[r,k]|^2. \quad \textbf{2.10}$$

2.2 Acoustic propagation

2.2.1 Microphones in free field

Suppose that M microphones are located in a free acoustic field and only one far-field source $s(t)$ is present i.e., one point source with plane wave propagation. With each ith microphone located at a corresponding location $\boldsymbol{r}_i$, the pressure waves at the microphones are given by:

$$x_i(t) = \alpha_i(\theta, \phi)s(t - \tau_i), i \in [1, M], \quad \textbf{2.11}$$

where the time delay at the i[th] microphone is

$$\tau_i = \frac{\boldsymbol{a}^T \boldsymbol{r}_i}{c}, \quad \textbf{2.12}$$

with c being the speed of sound, $\boldsymbol{a}$ is a unit vector that is oriented in the direction of propagation of the plane waves that carries the signal, $\alpha_i(\theta, \phi)$ is the attenuation factor caused by the directivity of the i[th] microphone, θ is the elevation angle and ϕ is the azimuth angle. The elevation and azimuth angle are respectively shown in blue and red in the following figure:

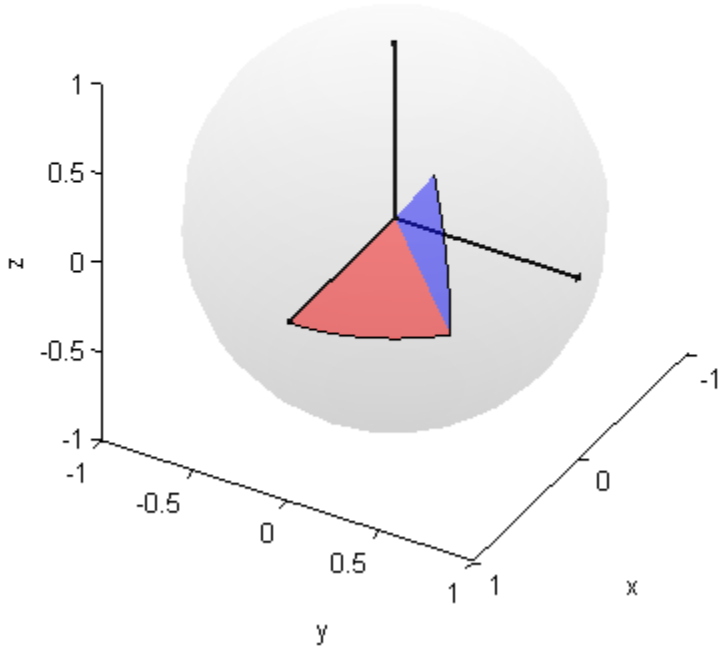

Figure 1: Elevation (θ) and azimuth (ϕ) angles

Thus the delay τ_i and the attenuation factor $\alpha_i(\theta, \phi)$ carry information about the direction of the source. The Fourier transform of the pressure waves at the microphones can be written as

$$X_i(\omega) = \alpha_i(\theta,\phi)e^{-\jmath\omega\tau_i}S(\omega),$$

or,

$$\boldsymbol{X}(\omega) = S(\omega)\boldsymbol{H}(\omega,\theta,\phi), \tag{2.13}$$

where

$$\boldsymbol{H}(\omega,\theta,\phi) = [\alpha_1(\theta,\phi)e^{-\jmath\omega\tau_1} \quad \alpha_2(\theta,\phi)e^{-\jmath\omega\tau_2} \quad \cdots \quad \alpha_M(\theta,\phi)e^{-\jmath\omega\tau_M}]^T,$$
$$\boldsymbol{a} = \boldsymbol{a}(\theta,\phi) = -[\sin(\theta)\cos(\phi) \quad \sin(\theta)\sin(\phi) \quad \cos(\theta)]^T.$$

Assuming that the signal $s(t)$ is a zero mean stationary ergodic process and has an autocorrelation of $\gamma_{ss}(t)$,

$$\begin{aligned}\mathbf{R}_{xx}(\tau) &= \mathcal{E}\big(\boldsymbol{x}(t)\boldsymbol{x}^T(t-\tau)\big)\\ &= \mathcal{E}\left(\left[\alpha_i(\theta,\phi)\alpha_j(\theta,\phi)s(t-\tau_i)s\big(t-\tau-\tau_j\big)\right]_{i,j\in[1,M]}\right)\\ &= \left[\alpha_i(\theta,\phi)\alpha_j(\theta,\phi)\gamma_{ss}(\tau-\big(\tau_i-\tau_j\big))\right]_{i,j\in[1,M]}\\ &= \left[\alpha_i(\theta,\phi)\alpha_j(\theta,\phi)\gamma_{ss}(\tau-\Delta\tau_{ij})\right]_{i,j\in[1,M]}.\end{aligned} \tag{2.14}$$

Note that operator $\boldsymbol{v}^T$ denotes the transpose of a vector. In the Fourier domain,

$$\begin{aligned}\boldsymbol{\Gamma}_{xx}(\omega,\theta,\phi) &= \Gamma_{ss}(\omega)\boldsymbol{H}(\omega,\theta,\phi)\boldsymbol{H}^H(\omega,\theta,\phi)\\ &= \Gamma_{ss}(\omega)\left[\alpha_n(\theta,\phi)\alpha_m(\theta,\phi)e^{-\frac{\jmath\omega}{c}(\Delta\boldsymbol{r}_{nm}^T\boldsymbol{a})}\right]_{n,m\in[1,M]},\end{aligned} \tag{2.15}$$

with

$$\Delta\boldsymbol{r}_{nm} = \boldsymbol{r}_n - \boldsymbol{r}_m,$$

and where $\boldsymbol{v}^H$ denotes the Hermitian transpose of the vector $\boldsymbol{v}$. If it is known how the directivity pattern behaves for a given source direction and if the positions of the microphones are known, then we can solve or approximate (θ,ϕ) from Eq.(2.15) if there is a solution. It is possible to obtain some information about (θ,ϕ) from the PSD matrix of the source $\boldsymbol{\Gamma}_{xx}(\omega,\theta,\phi)$, however estimating (θ,ϕ) from $\boldsymbol{\Gamma}_{xx}(\omega,\theta,\phi)$ is not the aim of this thesis.

2.2.2 The head related transfer function

Suppose that there a signal emitted by a source that is sampled at the eardrum of the left and right ear denoted by $x_l[n]$ and $x_r[n]$ respectively. The signal measured at each ear can be modeled as convolution (denoted by "*") of the transmitted signal by the head related impulse response (HRIR) of each respective ear or,

$$x_l[n] = h_l[n] * s[n]$$
$$x_r[n] = h_r[n] * s[n]$$

equivalently, **2.16**

$$\boldsymbol{x}[n] = \begin{bmatrix}h_l[n]\\ h_r[n]\end{bmatrix} * s[n] = \boldsymbol{h}[n] * s[n].$$

The discrete functions $h_l[n]$ and $h_r[n]$ are the left and right HRIR respectively. In the discrete time Fourier transform domain, the HRIR becomes the head-related frequency responses or, more commonly, head-related transfer functions (HRTFs). The HRTF is thus the ratio of the spectrum of the signal received at the ears over the source signal spectrum,

$$\frac{X_l(\omega)}{S(\omega)} = H_l(\omega),$$
$$\frac{X_r(\omega)}{S(\omega)} = H_r(\omega), \qquad \textbf{2.17}$$

or,

$$\frac{\boldsymbol{X}(\omega)}{S(\omega)} = \begin{bmatrix} H_l(\omega) \\ H_r(\omega) \end{bmatrix} = \boldsymbol{H}(\omega).$$

The diffraction, reflections and scattering information given by the HRTFs provides information about how the signal propagates in a given acoustic environment, effectively giving information about the depth of the source. The HRIRs depend on three spatial variables (r, ϕ, θ) being respectively the distance from the source to the eardrum, the elevation and the azimuth. Humans get information from the azimuth angle ϕ through the interaural time and level differences (ITD, ILD) and the elevation θ is estimated by the use of spectral coloring due to the pinnae [9]. Examples of estimation of ITDs and ILDs can be found in [10, 11]. The HRTFs vary from one individual to another, thus if we measure the HRTFs of an individual then the synthesis of a binaural signal with those HRTFs will be optimal only for this individual (although the synthesis may be acceptable for other individuals as well). It is possible for individuals to relearn sound localization when their HRTFs are modified [9]. Synthetizing a signal from HRTFs that were measured in an anechoic room will make the sound seem less natural and remove depth perception, hence the spatial variable r depends on the reverberation time of the impulse responses. The use of HRTFs is not the only method of recreating a binaural signal. A second way of creating a binaural signal is by recording live sound with microphones in a dummy head [12]. Binaural recordings preserve dynamical localization cues and can reproduce the effects of room reverberations that help to perceive the depth of the source. But this second method is very limited since it can only reproduce the recorded sound source, while the HRTF synthesis approach can be used to produce binaural signals for any sound source.

Let us proceed in deriving the PSD of the sampled signals at the left and right ear:

$$\begin{aligned} &\mathcal{E}(\boldsymbol{x}[m]\boldsymbol{x}^T[m-n]) \\ &= \mathcal{E}\begin{bmatrix} h_l[m] * s[m] \\ h_r[m] * s[m] \end{bmatrix}\begin{bmatrix} h_l[m-n] * s[m-n] \\ h_r[m-n] * s[m-n] \end{bmatrix}^T \\ &= \begin{bmatrix} \gamma_s[n] * h_l[n] * h_l[-n] & \gamma_s[n] * h_l[n] * h_r[-n] \\ \gamma_s[n] * h_r[n] * h_l[-n] & \gamma_s[n] * h_r[n] * h_r[-n] \end{bmatrix} \end{aligned} \qquad \textbf{2.18}$$

or,

$$= \gamma_s[n] * \boldsymbol{h}[n] * \boldsymbol{h}^T[-n]$$

The Fourier transform of the above equation gives,

$$\boldsymbol{\Gamma}_x(\omega) = \Gamma_s(\omega)\begin{bmatrix} |H_l(\omega)|^2 & H_l(\omega)H_r^*(\omega) \\ H_l^*(\omega)H_r(\omega) & |H_r(\omega)|^2 \end{bmatrix} \quad \textbf{2.19}$$

or,

$$= \Gamma_s(\omega)\boldsymbol{H}(\omega)\boldsymbol{H}^H(\omega).$$

2.2.3 Multiple input multiple output (MIMO) model for microphone arrays

A MIMO system for an array of *M* microphones and *N* acoustic sources has the following form at time *n:*

$$y_i[n] = x_i[n] + \eta_i[n]\,, i \in [1, M],$$

$$x_i[n] = \sum_{j=1}^{N} h_{i,j}[n] * s_j[n],$$

or in vector notation,

$$\boldsymbol{y}[n] = \boldsymbol{h}_{M\times N}[n] * \boldsymbol{s}[n] + \boldsymbol{\eta}[n] = \boldsymbol{x}[n] + \boldsymbol{\eta}[n],$$

where

$$\boldsymbol{h}_{M\times N}[n] = \begin{bmatrix} h_{1,1}[n] & h_{1,2}[n] & \cdots & h_{1,N}[n] \\ h_{2,1}[n] & h_{2,2}[n] & \cdots & h_{2,N}[n] \\ \vdots & \vdots & \ddots & \vdots \\ h_{M,1}[n] & h_{M,2}[n] & \cdots & h_{M,N}[n] \end{bmatrix}, \quad \textbf{2.20}$$

$$\boldsymbol{s}[n] = [s_1[n] \quad s_2[n] \quad \cdots \quad s_N[n]]^T,$$

$$\boldsymbol{y}[n] = [y_1[n] \quad y_2[n] \quad \cdots \quad y_M[n]]^T,$$

$$\boldsymbol{\eta}[n] = [\eta_1[n] \quad \eta_2[n] \quad \cdots \quad \eta_M[n]]^T,$$

with $h_{i,j}$ denoting the impulse response from the j[th] source to the i[th] microphone, η_i denotes the noise (diffuse, background, sensor, etc.) at the i[th] microphone , and y_i denotes the sampled signals at the i[th] microphone. The matrix $\boldsymbol{h}_{M\times N}[n]$ is thus an impulse response matrix, $\boldsymbol{s}[n]$ is the sources vector and $\boldsymbol{\eta}[n]$ is the diffuse noise vector. In the discrete time Fourier transform (DTFT) domain, we have the following:

$$Y_i(\omega) = \sum_{j=1}^{N} H_{i,j}(\omega)S_j(\omega) + N_i(\omega), i \in [1, M],$$

or in vector notation, **2.21**

$$\boldsymbol{Y}(\omega) = \boldsymbol{H}_{M\times N}(\omega)\boldsymbol{S}(\omega) + \boldsymbol{N}(\omega),$$

$$\boldsymbol{X}(\omega) = \boldsymbol{H}_{M\times N}(\omega)\boldsymbol{S}(\omega),$$

where

$$\boldsymbol{H}_{M\times N}(\omega)=\begin{bmatrix} H_{1,1}(\omega) & H_{1,2}(\omega) & \cdots & H_{1,N}(\omega) \\ H_{2,1}(\omega) & H_{2,2}(\omega) & \cdots & H_{2,N}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M,1}(\omega) & H_{M,2}(\omega) & \cdots & H_{M,N}(\omega) \end{bmatrix},$$

$$\boldsymbol{S}(\omega)=[S_1(\omega) \quad S_2(\omega) \quad \cdots \quad S_N(\omega)]^T,$$

$$\boldsymbol{Y}(\omega)=[Y_1(\omega) \quad Y_2(\omega) \quad \cdots \quad Y_M(\omega)]^T,$$

$$\boldsymbol{N}(\omega)=[N_1(\omega) \quad N_2(\omega) \quad \cdots \quad N_M(\omega)]^T.$$

The matrix $\boldsymbol{H}_{M\times N}(\omega)$ is a transfer function matrix, the vector $\boldsymbol{S}(\omega)$ is the sources spectrum vector, $\boldsymbol{X}(\omega)$ is the sources spectrum measured at the microphones, $\boldsymbol{N}(\omega)$ is the noise spectrum vector and $\boldsymbol{Y}(\omega)$ is the noisy sources spectrum. Assuming that the noise and the signals are uncorrelated, we get the following PSD matrix:

$$\boldsymbol{\Gamma}_y(\omega)=\boldsymbol{\Gamma}_{\boldsymbol{x}}(\omega)+\boldsymbol{\Gamma}_\eta(\omega),$$

with

$$\boldsymbol{\Gamma}_{\boldsymbol{x}}(\omega)=\boldsymbol{H}_{M\times N}(\omega)\boldsymbol{\Gamma}_s(\omega)\boldsymbol{H}^H_{M\times N}(\omega). \quad \textbf{2.22}$$

The matrix $\boldsymbol{\Gamma}_y(\omega)$ is the PSD matrix of the noisy signals $\boldsymbol{y}[n]$, $\boldsymbol{\Gamma}_{\boldsymbol{x}}(\omega)$ is the PSD matrix of the measured sources, $\boldsymbol{\Gamma}_s(\omega)$ is the PSD matrix at the sources , and $\boldsymbol{\Gamma}_\eta(\omega)$ is the PSD matrix of the diffuse noise.

2.3 Spatial coherence and correlation functions for isotropic sound fields

2.3.1 Spherical isotropic field

Now let's assume that an isotropic sound field measured at an array of *M* microphones is given by the sum over all angles of components with the following form:

$$\eta_i(t,\theta,\phi)=h_i(t,\theta,\phi)*\nu(t). \quad \textbf{2.23}$$

The impulse response $h_i(t,\theta,\phi)$ is associated to the point on the sphere from angles (θ,ϕ) to the i^{th} microphone, $\nu(t)$ is the noise pulsating from that point, and $\eta_i(t,\theta,\phi)$ is the noise measured at the i^{th} microphone caused by the point at angles (θ,ϕ).

Recall the result for a plane wave in Eq.(2.15),

$$\boldsymbol{\Gamma}_\eta(\omega,\theta,\phi)=\Gamma_\nu(\omega)\boldsymbol{H}(\omega,\theta,\phi)\boldsymbol{H}^H(\omega,\theta,\phi), \quad \textbf{2.24}$$

where $\boldsymbol{H}(\omega,\theta,\phi)$ is the transfer function vector, $\Gamma_\nu(\omega)$ is the PSD of the noise source.

The spatial coherence functions can be derived from averaging the PSD result for a plane wave in Eq.(2.24) over all possible plane waves that radiate from a surface surrounding the array. In the case of a spherically isotropic sound field, the surface is a sphere. The following integration follows,

$$\begin{aligned}\boldsymbol{\Gamma}_{\eta}(\omega) &= \frac{1}{4\pi}\int_{0}^{2\pi}\int_{0}^{\pi}\boldsymbol{\Gamma}_{\eta}(\omega,\theta,\phi)\sin(\theta)d\theta d\phi \\ &= \frac{\Gamma_{v}(\omega)}{4\pi}\int_{0}^{2\pi}\int_{0}^{\pi}\boldsymbol{H}(\omega,\theta,\phi)\boldsymbol{H}^{H}(\omega,\theta,\phi)\sin(\theta)d\theta d\phi \\ &= \Gamma_{v}(\omega)\boldsymbol{\Psi}(\omega)\end{aligned} \quad \textbf{2.25}$$

,where

$$\boldsymbol{\Gamma}_{\eta}(\omega) = \begin{bmatrix} \Gamma_{\eta_1}(\omega) & \Gamma_{\eta_1\eta_2}(\omega) & \cdots & \Gamma_{\eta_1\eta_M}(\omega) \\ \Gamma_{\eta_2\eta_1}(\omega) & \Gamma_{\eta_2}(\omega) & \cdots & \Gamma_{\eta_2\eta_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{\eta_M\eta_1}(\omega) & \Gamma_{\eta_M\eta_2}(\omega) & \cdots & \Gamma_{\eta_M}(\omega) \end{bmatrix}. \quad \textbf{2.26}$$

We call $\boldsymbol{\Psi}(\omega)$ the normalised spatial spectral matrix or the correlation[1] matrix given by:

$$\boldsymbol{\Psi}(\omega) = \frac{1}{4\pi}\int_{0}^{2\pi}\int_{0}^{\pi}\boldsymbol{H}(\omega,\theta,\phi)\boldsymbol{H}^{H}(\omega,\theta,\phi)\sin(\theta)d\theta d\phi. \quad \textbf{2.27}$$

The coherence function $\Phi_{\eta_i\eta_j}(\omega)$ between the ith and the jth microphone is defined as,

$$\Gamma_{\eta_i\eta_j}(\omega) = \sqrt{\Gamma_{\eta_i}(\omega)\Gamma_{\eta_j}(\omega)}\Phi_{\eta_i\eta_j}(\omega), \quad \textbf{2.28}$$

then $\boldsymbol{\Gamma}_{\eta}(\omega)$ can factorise in the following matrices:

$$\boldsymbol{\Gamma}_{\eta}(\omega) = \sqrt{\mathbf{D}(\omega)}\boldsymbol{\Phi}(\omega)\sqrt{\mathbf{D}(\omega)},$$

where

$$\mathbf{D}(\omega) = Diag\left(\Gamma_{\eta_1}(\omega), \Gamma_{\eta_2}(\omega), \dots, \Gamma_{\eta_M}(\omega)\right) = \boldsymbol{\Gamma}_{\eta}(\omega) \circ \boldsymbol{I},$$

and the positive definite matrix

$$\boldsymbol{\Phi}(\omega) = \begin{bmatrix} 1 & \Phi_{\eta_1\eta_2}(\omega) & \cdots & \Phi_{\eta_1\eta_M}(\omega) \\ \Phi_{\eta_2\eta_1}(\omega) & 1 & \cdots & \Phi_{\eta_2\eta_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{\eta_M\eta_1}(\omega) & \Phi_{\eta_M\eta_2}(\omega) & \cdots & 1 \end{bmatrix}. \quad \textbf{2.29}$$

where " $\circ$ " is the Schur product and $\sqrt{\mathbf{A}}$ is the principal square root of the matrix $\boldsymbol{A}$ [13].

[1] In the present work, we make a distinction between correlation and coherence. But it happens in some cases that they are effectively equal.

Decomposing the right hand side of Eq.(2.26) in the same way as we just did in Eq.(2.29) gives a formula for the PSDs and the coherence functions at each microphone in terms of the transfer functions. The link between the normalised spatial spectral matrix and the coherence matrix is straightforward, it is given by:

$$\boldsymbol{\Psi}(\omega) = \sqrt{\boldsymbol{\Psi}(\omega) \circ \boldsymbol{I}}\boldsymbol{\Phi}(\omega)\sqrt{\boldsymbol{\Psi}(\omega) \circ \boldsymbol{I}}. \tag{2.30}$$

Another relation is,

$$\Gamma_v(\omega)\boldsymbol{\Psi}(\omega) = \sqrt{\mathbf{D}(\omega)}\boldsymbol{\Phi}(\omega)\sqrt{\mathbf{D}(\omega)}. \tag{2.31}$$

From the above formula we get by looking at the diagonals

$$\Gamma_{\eta_i}(\omega) = \frac{\Gamma_v(\omega)}{4\pi}\int_0^{2\pi}\int_0^{\pi} |H_i(\omega,\theta,\phi)|^2 \sin(\theta) d\theta d\phi, \tag{2.32}$$

and looking at the off diagonal terms:

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{\int_0^{2\pi}\int_0^{\pi} H_i(\omega,\theta,\phi)H_j^*(\omega,\theta,\phi)\sin(\theta)d\theta d\phi}{\sqrt{\int_0^{2\pi}\int_0^{\pi}|H_i(\omega,\theta,\phi)|^2 \sin(\theta)d\theta d\phi \int_0^{2\pi}\int_0^{\pi}\left|H_j(\omega,\theta,\phi)\right|^2 \sin(\theta)d\theta d\phi}}. \tag{2.33}$$

If the following assumption is made,

$$\Gamma_{\eta_i}(\omega) = \Gamma_{\eta_j}(\omega) = \Gamma_{\eta}(\omega), \tag{2.34}$$

then it is implied that

$$\int_0^{2\pi}\int_0^{\pi} |H_i(\omega,\theta,\phi)|^2 \sin(\theta)d\theta d\phi = \int_0^{2\pi}\int_0^{\pi} \left|H_j(\omega,\theta,\phi)\right|^2 \sin(\theta)d\theta d\phi. \tag{2.35}$$

If ever Eq.(2.35) doesn't hold, it is simpler to work with $\boldsymbol{\Psi}(\omega)$ instead since it is mathematically easier to work with the left hand side of Eq.(2.31).

2.3.1.1 Example in free field with isotropic directivity sensor pattern

If we have a free field with isotropic patterns, the measured sound at an array of microphones is

$$\eta_i(t) = v\left(t - \frac{\boldsymbol{a}^T\boldsymbol{r}_i}{c}\right). \tag{2.36}$$

where $\boldsymbol{a}^T = -[\sin(\theta)\cos(\phi) \quad \sin(\theta)\sin(\phi) \quad \cos(\theta)]^T$.

The expression for the coherence using Eq.(2.33) is:

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} e^{-\jmath\omega\frac{\boldsymbol{a}^T(\boldsymbol{r}_i-\boldsymbol{r}_j)}{c}}\sin(\theta)d\theta d\phi. \quad \textbf{2.37}$$

Without loss of generality, let's assume that the sensors are located on the z axis. Then the integral can be evaluated as

$$\begin{aligned}\Phi_{\eta_i\eta_j}(\omega) &= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\exp\left(\jmath\omega\frac{|\Delta\boldsymbol{r}_{ij}|}{c}\cos(\theta)\right)\sin(\theta)d\theta d\phi \\ &= \frac{\sin\left(\omega\frac{|\Delta\boldsymbol{r}_{ij}|}{c}\right)}{\omega\frac{|\Delta\boldsymbol{r}_{ij}|}{c}}.\end{aligned} \quad \textbf{2.38}$$

The equation above is valid for any difference vector $\Delta\boldsymbol{r}_{ij}$, due to symmetry. In this case,

$$\Gamma_\eta(\omega) = \Gamma_v(\omega), \quad \textbf{2.39}$$

and the coherence matrix equals the normalised spatial spectral matrix(or correlation matrix):

$$\boldsymbol{\Phi}(\omega) = \boldsymbol{\Psi}(\omega). \quad \textbf{2.40}$$

2.3.1.2 Example with a rigid sphere and isotropic sensors(★)

To model the effects of scattering due to the presence of the head, the head is modeled as a sphere of radius a. Due to the scattering effects of the sphere, there will be outgoing waves that bounce off the sphere. The total field will be the sum of both the incoming waves and outgoing waves. In the Fourier domain, the sound field would be modeled as

$$N(\omega,\theta,\phi,\theta_i,\phi_i) = p_{in}(\omega,r,\theta,\phi,\theta_i,\phi_i) + p_{out}(\omega,r,\theta,\phi,\theta_i,\phi_i), \quad \textbf{2.41}$$

where $p_{in}(\omega,r,\theta,\phi,\theta_i,\phi_i)$, $p_{out}(\omega,r,\theta,\phi,\theta_i,\phi_i)$ are the incoming and outgoing pressure waves from angles (θ,ϕ) to the microphone angles (θ_i,ϕ_i), $N(\omega,\theta,\phi,\theta_i,\phi_i)$ is the noise spectrum pressure caused by the point on the surface at angles (θ,ϕ).

The incident field term, assuming that the sensors are in the z axis, is found to be

$$p_{in}(\omega,r,\theta,\phi,\theta_i,\phi_i) = V(\omega)\exp(\jmath\boldsymbol{k}^T\,\boldsymbol{r}_i), \quad \textbf{2.42}$$

where $\boldsymbol{k} = -\frac{\omega}{c}\boldsymbol{a}$ is the wavenumber vector, and $V(\omega)$ is the spectrum of sound pulsating out of the point from angles (θ,ϕ). The spherical geometry permits a convenient expansion in terms of basis functions that can be found by solving the Helmholtz equation. In the chapter 6 of [14], it is shown that the incoming waves term coming from the angle (θ,ϕ) are given by:

$$p_{in}(\omega, r, \theta, \phi, \theta_i, \phi_i) = V(\omega)4\pi \sum_{n=0}^{\infty} j_n\left(\omega\frac{r}{c}\right)\jmath^n \sum_{m=-n}^{n} Y_n^m(\theta_i, \phi_i)Y_n^{*m}(\theta, \phi), \quad \textbf{2.43}$$

where j_n is the spherical Bessel function of order *n*. $Y_n^m(\theta, \phi)$ are the spherical harmonics of order n defined by:

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos(\theta))e^{\jmath m\phi}, \quad \textbf{2.44}$$

with P_n^m being the associated Legendre function. The sum of the spherical harmonics can be expanded as

$$\sum_{m=-n}^{n} Y_n^m(\theta_i, \phi_i)Y_n^{*m}(\theta, \phi) = \sum_{m=-n}^{n} \frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}P_n^m(\cos(\theta))P_n^m(\cos(\theta_i))e^{-\jmath m(\phi-\phi_i)}. \quad \textbf{2.45}$$

Since

$$P_n^{-m}(x) = (-1)^m\frac{(n-m)!}{(n+m)!}P_n^m(x), \quad \textbf{2.46}$$

the Eq.(2.45) becomes

$$\begin{aligned}
&\sum_{m=-n}^{n} Y_n^m(\theta_i, \phi_i)Y_n^{*m}(\theta, \phi) \\
&= \frac{2n+1}{4\pi}P_n(\cos(\theta))P_n(\cos(\theta_i)) \\
&\quad + \sum_{m=1}^{n} \frac{2n+1}{4\pi}\left(\frac{(n-m)!}{(n+m)!}P_n^m(\cos(\theta))P_n^m(\cos(\theta_i))e^{-\jmath m(\phi-\phi_i)}\right. \\
&\quad \left. + \frac{(n+m)!}{(n-m)!}P_n^{-m}(\cos(\theta))P_n^{-m}(\cos(\theta_i))e^{\jmath m(\phi-\phi_i)}\right) \\
&= \frac{2n+1}{4\pi}\left[P_n(\cos(\theta))P_n(\cos(\theta_i))\right. \\
&\quad \left. + 2\sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!}P_n^m(\cos(\theta))P_n^m(\cos(\theta_i))\cos\big(m(\phi-\phi_i)\big)\right].
\end{aligned} \quad \textbf{2.47}$$

In [15] the addition theorem of the spherical harmonics is used [16] which states that the sum of spherical harmonics can be reduced to a single Legendre function so the above equation becomes:

$$\sum_{m=-n}^{n} Y_n^m(\theta,\phi)Y_n^{*m}(\theta_i,\phi_i) = \frac{2n+1}{4\pi}P_n(\cos(\theta)\cos(\theta_i)+\sin(\theta)\sin(\theta_i)\cos(\phi-\phi_i)). \quad \textbf{2.48}$$

Also in [15] a convenient notation for the argument of the Legendre function is used i.e.,

$$\Theta_{\mathrm{i}} = \cos(\theta)\cos(\theta_i)+\sin(\theta)\sin(\theta_i)\cos(\phi-\phi_i). \quad \textbf{2.49}$$

Substituting the above result into Eq.(2.43) we get a simpler expression for p_{in} that the angular dependency is reduced to the quantity Θ_{i}:

$$p_{in}(\omega,r,\Theta_{\mathrm{i}}) = V(\omega)\sum_{n=0}^{\infty} j_n\left(\omega\frac{r}{c}\right)(\mathrm{J})^n(2n+1)P_n(\Theta_{\mathrm{i}}). \quad \textbf{2.50}$$

For a rigid sphere the boundary condition states that the radial velocity vanishes on the sphere, that is

$$\frac{\partial}{\partial r}\left(p_{in}(\omega,r,\Theta_{\mathrm{i}})+p_{out}(\omega,r,\Theta_{\mathrm{i}})\right)=0. \quad \textbf{2.51}$$

The scattered field is expanded in outgoing waves [14]

$$p_{out}(\omega,r,\Theta_{\mathrm{i}}) = \sum_{n=0}^{\infty} C_n(\omega)h_n^{(1)}\left(\omega\frac{r}{c}\right)P_n(\Theta_{\mathrm{i}}), \quad \textbf{2.52}$$

where $h_n^{(1)}$ is a spherical Hankel function and it will be written h_n for short. The constant $C_n(\omega)$ is an unknown to be determined with inserting Eq.(2.52) and (2.50) into (2.51) we get:

$$C_n(\omega) = -V(\omega)\frac{j_n'\left(\omega\frac{a}{c}\right)}{h_n'\left(\omega\frac{a}{c}\right)}\mathrm{J}^n(2n+1). \quad \textbf{2.53}$$

So the total field is given by

$$p_{tot}(\omega,r,\Theta_{\mathrm{i}}) = p_{in}(\omega,r,\Theta_{\mathrm{i}})+p_{out}(\omega,r,\Theta_{\mathrm{i}})$$
$$= V(\omega)\sum_{n=0}^{\infty}\left(j_n\left(\omega\frac{r}{c}\right)-\frac{j_n'\left(\omega\frac{a}{c}\right)}{h_n'\left(\omega\frac{a}{c}\right)}h_n\left(\omega\frac{r}{c}\right)\right)(\mathrm{J})^n(2n+1)P_n(\Theta_{\mathrm{i}}). \quad \textbf{2.54}$$

In[15] a constant $b_n\left(\omega\frac{r}{c}\right)$ is defined to further compress the notation. We present a version of it without the factor 4π found in[15] and it is defined by:

$$b_n\left(\omega\frac{r}{c}\right) = \quad j_n\left(\omega\frac{r}{c}\right) - \frac{j_n'\left(\omega\frac{a}{c}\right)}{h_n'\left(\omega\frac{a}{c}\right)} h_n\left(\omega\frac{r}{c}\right); rigid\ sphere$$
$$b_n\left(\omega\frac{r}{c}\right) = \quad j_n\left(\omega\frac{r}{c}\right); free\ field \tag{2.55}$$

The coherence[2] between any two points is

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} p(\omega, r_i, \Theta_{\mathrm{i}})\, p^*\left(\omega, r_j, \Theta_j\right)\sin(\theta)d\theta d\phi. \tag{2.56}$$

To evaluate the above integral we use the following integral relation

$$\begin{aligned}&\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} P_n(\Theta_{\mathrm{i}})\, P_m\left(\Theta_{\mathrm{j}}\right)\sin(\theta)d\theta d\phi \\ &= \frac{1}{4\pi}P_n\left(\Theta_{\mathrm{ij}}\right)\int_0^{2\pi}\int_0^{\pi} P_n(\cos(\theta))\, P_m(\cos(\theta))\sin(\theta)d\theta d\phi \\ &= \frac{P_n\left(\Theta_{\mathrm{ij}}\right)}{2n+1}\delta(n-m).\end{aligned} \tag{2.57}$$

We used in the integral the quantity[3]

$$\Theta_{\mathrm{ij}} = \cos\left(\theta_j\right)\cos(\theta_i) + \sin\left(\theta_j\right)\sin(\theta_i)\cos\left(\phi_i - \phi_j\right). \tag{2.58}$$

Inserting Eq.(2.58), (2.57),(2.56) and (2.55) into (2.56) we get

$$\Phi_{\eta_i\eta_j}(\omega) = \sum_{n=0}^{\infty} b_n\left(\frac{\omega}{c}r_i\right) b_n^*\left(\frac{\omega}{c}r_j\right)(2n+1)P_n\left(\Theta_{\mathrm{ij}}\right). \tag{2.59}$$

Surprisingly, $\Phi_{\eta_i\eta_j}(\omega)$ is real[4] for whatever positive value of r_i and r_j. It is interesting to note that in free field

[2] In this case the coherence is equivalent to the correlation. This can be verified numerically.

[3] This is the angle between the ith and jth point.

[4] This was verified numerically.

$$\Phi_{\eta_i\eta_j}(\omega) = \sum_{n=0}^{\infty} j_n\left(\frac{\omega}{c}r_i\right) j_n\left(\frac{\omega}{c}r_j\right)(2n+1)P_n\left(\Theta_{\mathrm{ij}}\right) = \frac{\sin\left(\omega\frac{|\Delta \boldsymbol{r}_{ij}|}{c}\right)}{\omega\frac{|\Delta \boldsymbol{r}_{ij}|}{c}}. \quad \textbf{2.60}$$

If both sensors have the same distance from the origin, then

$$\Phi_{\eta_i\eta_j}(\omega) = \sum_{n=0}^{\infty} j_n^2\left(\frac{\omega}{c}r\right)(2n+1)P_n\left(\Theta_{\mathrm{ij}}\right) = \frac{\sin\left(\omega\frac{r}{c}\sqrt{2\left(1-\Theta_{\mathrm{ij}}\right)}\right)}{\omega\frac{r}{c}\sqrt{2\left(1-\Theta_{\mathrm{ij}}\right)}}. \quad \textbf{2.61}$$

If $\Theta_{\mathrm{ij}} = -1$, that is if sensors are located at opposite ends of an open sphere, then

$$\Phi_{\eta_i\eta_j}(\omega) = \sum_{n=0}^{\infty} j_n^2\left(\frac{\omega}{c}r\right)(2n+1)(-1)^n = \frac{\sin\left(\omega\frac{2r}{c}\right)}{\omega\frac{2r}{c}} = \frac{\sin\left(\frac{\omega d}{c}\right)}{\omega\frac{d}{c}}, \quad \textbf{2.62}$$

where d is the diameter of the sphere.

2.3.2 Cylindrical isotropic sound field measured in free field

The following PSD matrix can be derived by averaging the plane wave PSD matrices over a cylinder that irradiates the sound

$$\boldsymbol{\Gamma}_{\eta}(\omega) = \frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{\Gamma}_{\eta}(\omega,\theta,\phi)\,d\theta = \frac{\Gamma_v(\omega)}{2\pi}\int_0^{2\pi} \boldsymbol{H}(\omega,\theta,\phi)\boldsymbol{H}^H(\omega,\theta,\phi)\,d\theta. \quad \textbf{2.63}$$

Following the same line of thought that allowed us to find an expression for the coherence functions in terms of the transfer functions in the case of a spherical sound field we get:

$$\Gamma_{\eta_i}(\omega) = \frac{\Gamma_v(\omega)}{2\pi}\int_0^{2\pi} |H_i(\omega,\theta,\phi)|^{\mathbf{2}}\,d\theta. \quad \textbf{2.64}$$

Here is the expression for the spatial coherence in terms of the directivity patterns and the distance between the corresponding microphones:

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{\int_0^{2\pi} H_i(\omega,\theta,\phi)H_j^*(\omega,\theta,\phi)\,d\theta}{\sqrt{\int_0^{2\pi}|H_i(\omega,\theta,\phi)|^2 d\theta \int_0^{2\pi}\left|H_j(\omega,\theta,\phi)\right|^2 d\theta}}. \quad \textbf{2.65}$$

2.3.2.1 Example in free field

Now let's assume that a cylindrically isotropic sound field is measured at an array of M microphones in free field such as:

$$\eta_i(t) = \alpha_i(\theta,\phi)v\left(t - \frac{\boldsymbol{a}^T\boldsymbol{r}_i}{c}\right), \quad \textbf{2.66}$$

where $\boldsymbol{a}^T = -[\cos(\theta) \quad \sin(\theta) \quad 0]$. The directivity pattern of the ith microphone is $\alpha_i(\theta,\phi)$.

The expression for the coherence functions is:

$$\Gamma_{\eta_i}(\omega) = \frac{\Gamma_v(\omega)}{2\pi}\int_0^{2\pi} \alpha_i^2(\theta,\phi)\,d\theta. \quad \textbf{2.67}$$

If the amount of energy of each directivity pattern is the same, then the sound PSDs are the same at each microphone. Here is the expression for the coherence in terms of the directivity patterns and the distance between the corresponding microphones:

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{\int_0^{2\pi} \alpha_i(\theta,\phi)\alpha_j(\theta,\phi)\,e^{-\jmath\omega\frac{\boldsymbol{a}^T(\boldsymbol{r_i}-\boldsymbol{r_j})}{c}}d\theta}{\sqrt{\int_0^{2\pi}\alpha_i^2(\theta,\phi)\int_0^{2\pi}\alpha_j^2(\theta,\phi)}}. \quad \textbf{2.68}$$

A more general equation is given in [17] that is shown to be:

$$\Phi_{\eta_1\eta_2}(\omega) = \frac{\mathcal{E}\left[T_1(\theta,\phi,\omega)T_2^*(\theta,\phi,\omega)e^{-\jmath\omega\frac{\boldsymbol{a}^T(\boldsymbol{r_i}-\boldsymbol{r_j})}{c}}\right]}{\sqrt{\mathcal{E}[|T_1(\theta,\phi,\omega)|^2]\mathcal{E}[|T_2(\theta,\phi,\omega)|^2]}}. \quad \textbf{2.69}$$

where $\mathcal{E}[\text{" "}]$ in this section is understood to be the average over the domain of (θ,ϕ).

If the directivity patterns are isotropic ($\alpha_i(\theta,\phi) = 1$), then we have

$$\Gamma_{\eta_i\eta_j}(\omega) = \frac{\Gamma_v(\omega)}{2\pi}\int_0^{2\pi} e^{-\jmath\omega\frac{\boldsymbol{a}^T(\boldsymbol{r_i}-\boldsymbol{r_j})}{c}}d\theta. \quad \textbf{2.70}$$

If the propagation vector is given by $\boldsymbol{a}^T = -[\cos(\theta) \quad \sin(\theta) \quad 0]$, then in the integrand's argument $\boldsymbol{a}^T(\boldsymbol{r_i} - \boldsymbol{r_j})$ can be expanded as

$$\boldsymbol{a}^T(\boldsymbol{r_i}-\boldsymbol{r_j}) = \boldsymbol{a}^T\Delta\boldsymbol{r}_{ij} = \left(\Delta r_{ij_x}\cos(\theta) + \Delta r_{ij_y}\sin(\theta)\right) = \sqrt{\Delta r_{ij_x}^2 + \Delta r_{ij_y}^2}\cos(\theta+\psi). \quad \textbf{2.71}$$

It is noted that the coherence will be independent of the difference of the two microphones in the z axis. Suppose without loss of generality that the microphones are in the x-y plane, then we can write $\sqrt{\Delta r_x^2 + \Delta r_y^2} = \left|\Delta\boldsymbol{r}_{ij}\right|$. The integral will be independent of ψ. It can be rewritten as

$$\Gamma_{\eta_i\eta_j}(\omega) = \frac{\Gamma_v(\omega)}{2\pi}\int_0^{2\pi} e^{\jmath\omega\frac{|\Delta\boldsymbol{r}_{ij}|}{c}\cos(\theta)}\,d\theta = \Gamma_v(\omega)J_0\left(\omega\frac{\left|\Delta\boldsymbol{r}_{ij}\right|}{c}\right), \quad \textbf{2.72}$$

where J_0 is the Bessel function of the first kind of order 0. It follows that the coherences are

$$\Phi_{\eta_i\eta_j}(\omega) = J_0\left(\omega\frac{\left|\Delta\boldsymbol{r}_{ij}\right|}{c}\right). \quad \textbf{2.73}$$

It turns out that $\boldsymbol{\Psi}(\omega)$ equals the coherence matrix $\boldsymbol{\Phi}(\omega)$ in this case too.

2.3.3 Sound emanating from a general surface.

An integral of the form $\frac{1}{A}\int_{\boldsymbol{\Omega}\in S^2}$ " "$d\boldsymbol{\Omega}$ is done over the surface that emanates a sound, A is the total area of the surface and $d\boldsymbol{\Omega}$ is the differential on this surface [18]. By analogy, all preceding formulas can be rewritten in the form,

$$\boldsymbol{\Gamma}_\eta(\omega) = \frac{\Gamma_v(\omega)}{A}\int_{\boldsymbol{\Omega}\in S^2}\boldsymbol{H}(\omega,\boldsymbol{\Omega})\boldsymbol{H}^H(\omega,\boldsymbol{\Omega})d\boldsymbol{\Omega}. \quad \textbf{2.74}$$

The coherences are given by,

$$\Phi_{\eta_i\eta_j}(\omega) = \frac{\int_{\boldsymbol{\Omega}\in S^2} H_i(\omega,\boldsymbol{\Omega})H_j^*(\omega,\boldsymbol{\Omega})d\boldsymbol{\Omega}}{\sqrt{\int_{\boldsymbol{\Omega}\in S^2}|H_i(\omega,\boldsymbol{\Omega})|^2 d\boldsymbol{\Omega}\int_{\boldsymbol{\Omega}\in S^2}\left|H_j(\omega,\boldsymbol{\Omega})\right|^2 d\boldsymbol{\Omega}}}, \quad \textbf{2.75}$$

and the normalised spatial spectral matrix or correlation matrix is given by:

$$\boldsymbol{\Psi}(\omega) = \frac{1}{A}\int_{\boldsymbol{\Omega}\in S^2}\boldsymbol{H}(\omega,\boldsymbol{\Omega})\boldsymbol{H}^H(\omega,\boldsymbol{\Omega})d\boldsymbol{\Omega}. \quad \textbf{2.76}$$

Here is a brief recapitulation of the previous examples done in a free field with isotropic microphones.

Sound field provenance	General integral	Coherences
Spherical	$\frac{1}{A}\int_{\mathbf{\Omega}\in S^2} ""d\mathbf{\Omega} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} "" \sin(\theta)d\theta d\phi$	$\Phi_{\eta_i\eta_j}(\omega) = \frac{\sin\left(\omega\frac{\lvert\Delta r_{ij}\rvert}{c}\right)}{\omega\frac{\lvert\Delta r_{ij}\rvert}{c}}$
Cylindrical	$\frac{1}{A}\int_{\mathbf{\Omega}\in S^2} ""d\mathbf{\Omega} = \frac{1}{2\pi}\int_0^{2\pi} ""d\theta$	$\Phi_{\eta_i\eta_j}(\omega) = J_0\left(\omega\frac{\lvert\Delta r_{ij}\rvert}{c}\right)$

Table 1: Free field sound coherences with isotropic microphones

2.3.3.1 Integral discretization implementation of coherences

If an analytical model is not available it is still possible to approximate the coherences if we have at least a sufficient amount of transfer functions to approximate the integral in Eq.(2.75) with a discretized version of it. We can also discretize it if the integrands are known, but have no closed-form solution as done in [19]. For example, if the sound comes from a cylinder of infinite radius,

$$\begin{aligned}\Phi_{\eta_i\eta_j}(\omega) &= \frac{\int_0^{2\pi} H_i(\omega,\theta)H_j^*(\omega,\theta)d\theta}{\sqrt{\int_0^{2\pi}|H_i(\omega,\theta)|^2 d\theta \int_0^{2\pi}\left|H_j(\omega,\theta)\right|^2 d\theta}} \\ &\cong \frac{\sum_{k=0}^{K-1} H_i\left(\omega,\frac{2\pi k}{K}\right)H_j^*\left(\omega,\frac{2\pi k}{K}\right)}{\sqrt{\sum_{k=0}^{K-1}\left|H_i\left(\omega,\frac{2\pi k}{K}\right)\right|^2 \sum_{k=0}^{K-1}\left|H_j\left(\omega,\frac{2\pi k}{K}\right)\right|^2}}.\end{aligned} \quad \textbf{2.77}$$

The normalised spatial spectral matrix or correlation matrix $\mathbf{\Psi}$ terms are in the numerator of the last equation:

$$\begin{aligned}\Psi_{\eta_i\eta_j}(\omega) &= \frac{1}{2\pi}\int_0^{2\pi} H_i(\omega,\theta)H_j^*(\omega,\theta)d\theta \\ &\cong \frac{1}{K}\sum_{k=0}^{K-1} H_i\left(\omega,\frac{2\pi k}{K}\right)H_j^*\left(\omega,\frac{2\pi k}{K}\right).\end{aligned} \quad \textbf{2.78}$$

The matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are supposed to be well conditioned, that is, the ratio of the largest to smallest singular value is not too large.

2.3.4 Characteristics of the sound PSD matrix for isotropic sound fields

There is a relation between optimal non-causal Wiener filters and the sound spectrum at each microphone. The filter used to best approximate (or linearly predict) the i^{th} noise spectrum from the j^{th} noise spectrum is given by:

$$\mathcal{E}\left(N_i(\omega)|N_j(\omega)\right) = \widehat{N}_i(\omega) = \frac{\mathcal{E}\left(N_i(\omega)N_j^*(\omega)\right)}{\mathcal{E}\left(\left|N_j(\omega)\right|^2\right)} N_j(\omega)$$
$$= H_{W_{i,j}}(\omega)N_j(\omega)$$

where, **2.79**

$$H_{W_{i,j}}(\omega) = \sqrt{\frac{\Gamma_{\eta_i}(\omega)}{\Gamma_{\eta_j}(\omega)}}\Phi_{\eta_i\eta_j}(\omega) = \sqrt{\frac{\Psi_{\eta_i}(\omega)}{\Psi_{\eta_j}(\omega)}}\Phi_{\eta_i\eta_j}(\omega)$$
$$= \frac{\Psi_{\eta_i\eta_j}(\omega)}{\Psi_{\eta_j}(\omega)}.$$

If the sound PSD at the microphones are identical, i.e. $\Gamma_{\eta_i}(\omega) = \Gamma_{\eta_j}(\omega) = \Gamma_{\eta}(\omega)$ then,

$$H_{W_{i,j}}(\omega) = \Phi_{\eta_i\eta_j}(\omega). \quad \textbf{2.80}$$

Hence the Wiener filters are the spatial coherence functions in this special case that occurs whenever

$$\int_{\boldsymbol{\Omega}\in S^2} |H_i(\omega,\boldsymbol{\Omega})|^2\, d\boldsymbol{\Omega} = \int_{\boldsymbol{\Omega}\in S^2} \left|H_j(\omega,\boldsymbol{\Omega})\right|^2 d\boldsymbol{\Omega}. \quad \textbf{2.81}$$

Note that the equation above is a more general version of Eq.(2.35).

It should be noted that $\boldsymbol{\Gamma}_{\eta}(\omega)$ is Hermitian and almost certainly a positive definite matrix when estimated in practice, so it can be factored in the following form,

$$\boldsymbol{\Gamma}_{\eta}(\omega) = \mathbf{V}(\omega)\boldsymbol{\Lambda}(\omega)\mathbf{V}^{\mathrm{H}}(\omega), \quad \textbf{2.82}$$

where $\mathbf{V}(\omega)$ is a unitary matrix and $\boldsymbol{\Lambda}(\omega)$ is a positive diagonal matrix with positive eigenvalues. Since $\boldsymbol{\Gamma}_{\eta}(\omega)$ is positive definite it follows that $\boldsymbol{\Phi}(\omega)$ also is positive definite. The proof goes in the following way:

For any non-zero vector $\boldsymbol{z} \in \mathbb{C}^M$, $\boldsymbol{\Gamma}_{\eta}(\omega)$ is positive definite so $\boldsymbol{z}^H\boldsymbol{\Gamma}_{\eta}(\omega)\boldsymbol{z} > 0$.

But, $\boldsymbol{\Gamma}_{\eta}(\omega) = \sqrt{\mathbf{D}(\omega)}\boldsymbol{\Phi}(\omega)\sqrt{\mathbf{D}(\omega)}$ where $\mathbf{D}(\omega)$ is a positive definite diagonal matrix.

So

$$\boldsymbol{z}^H\boldsymbol{\Gamma}_{\eta}(\omega)\boldsymbol{z} = \boldsymbol{z}^H\sqrt{\mathbf{D}(\omega)}\boldsymbol{\Phi}(\omega)\sqrt{\mathbf{D}(\omega)}\boldsymbol{z} = \boldsymbol{w}^H\boldsymbol{\Phi}(\omega)\boldsymbol{w} > 0, \quad \textbf{2.83}$$

and

$$\boldsymbol{z}^H\boldsymbol{\Gamma}_{\eta}(\omega)\boldsymbol{z} = \Gamma_{v}(\omega)\boldsymbol{z}^H\boldsymbol{\Psi}(\omega)\boldsymbol{z} = \boldsymbol{u}^H\boldsymbol{\Psi}(\omega)\boldsymbol{u} > 0, \quad \textbf{2.84}$$

hence $\boldsymbol{\Phi}(\omega)$ and $\boldsymbol{\Psi}(\omega)$ are positive definite. This implies that the matrices have a principal square root [13],

$$\boldsymbol{\Phi}(\omega) = \boldsymbol{\Phi}^{1/2}(\omega)\boldsymbol{\Phi}^{1/2}(\omega). \quad \textbf{2.85}$$

From Eq.(2.30)

$$\begin{aligned}\boldsymbol{\Psi}(\omega) &= \boldsymbol{\Psi}^{1/2}(\omega)\boldsymbol{\Psi}^{1/2}(\omega)\\ &= \sqrt{\boldsymbol{\Psi}(\omega)\circ \boldsymbol{I}}\boldsymbol{\Phi}^{1/2}(\omega)\boldsymbol{\Phi}^{1/2}(\omega)\sqrt{\boldsymbol{\Psi}(\omega)\circ \boldsymbol{I}}.\end{aligned} \quad \textbf{2.86}$$

This means that the sound spectra at the microphones can be interpreted as a linear combination of non-causal filtered zero mean white wide sense stationary (WSS) Gaussian random variable.

$$\begin{aligned}\mathbf{N}(\omega) &= \sqrt{\Gamma_v(\omega)}\boldsymbol{\Psi}^{1/2}(\omega)\boldsymbol{\mathcal{W}}(\omega),\\ \boldsymbol{\mathcal{W}}(\omega) &= [\mathcal{W}_1(\omega) \quad \mathcal{W}_2(\omega) \quad \cdots \quad \mathcal{W}_{\mathrm{M}}(\omega)]^T,\\ \mathcal{E}\big(\boldsymbol{\mathcal{W}}(\omega)\big) &= 0, \mathcal{E}\left(\boldsymbol{\mathcal{W}}(\omega)\boldsymbol{\mathcal{W}}^{\mathrm{H}}(\omega)\right) = \boldsymbol{I}.\end{aligned} \quad \textbf{2.87}$$

It the case where $\Gamma_{\eta_i}(\omega) = \Gamma_{\eta_j}(\omega) = \Gamma_{\eta}(\omega)$, the above equation becomes

$$\mathbf{N}(\omega) = \sqrt{\Gamma_{\eta}(\omega)}\boldsymbol{\Phi}^{1/2}(\omega)\boldsymbol{\mathcal{W}}(\omega). \quad \textbf{2.88}$$

If the nth and mth sensors are very close to each other, $\boldsymbol{\Phi}(\omega)$ will likely become ill-conditioned since $\Phi_{\eta_i\eta_m}(\omega) \cong \Phi_{\eta_i\eta_n}(\omega) \cong \Phi_{\eta_n\eta_i}(\omega) \cong \Phi_{\eta_m\eta_i}(\omega)$. This will make the nth and mth rows and columns be very close numerically. If only a pair of rows is identical, the rank would be reduced by 1, so that the approximate rank of $\boldsymbol{\Phi}(\omega)$ becomes $M-1$.

2.4 Basic equations of the additive noise problem

In this section, the usual formulation of the additive noise problem is presented. Optimisation theory will be used, in order to derive filters that can reduce the noise and to show that we need to know the variance matrix of the noise (hence the motivation for this thesis). The notation presented here will be used for the remainder of the work.

Say we have measured the spectrogram of signals received at *M* different microphones from *N* sources with diffuse noise in the STFT domain with the MIMO perspective,

$$Y_i\,[r,k] = X_i[r,k] + N_i[r,k], i \in [1,M]$$

with

$$X_i[r,k] = \sum_{j=1}^{N} H_{i,j}[r,k]S_j[r,k]\,, i \in [1,M] \quad \textbf{2.89}$$

or

$$\boldsymbol{Y}\,[r,k] = \boldsymbol{X}[r,k] + \boldsymbol{N}[r,k],$$

with

$$\boldsymbol{X}[r,k] = \boldsymbol{H}_{M\times N}[r,k]\boldsymbol{S}[r,k].$$

Using the fact that the signal and the noise are uncorrelated we get the variance of the signal *y* at frequency-bin *k* and time-frame *r*. From now on, the frequency bin and the time frame won't be shown for brevity, unless stated otherwise. For example, the last equation will be written as

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N},$$

with

$$\boldsymbol{X} = \boldsymbol{H}_{M\times N}\boldsymbol{S}. \tag{2.90}$$

The PSD matrices are:

$$\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_x + \boldsymbol{\Gamma}_\eta,$$

where

$$\boldsymbol{\Gamma}_x = \boldsymbol{H}_{M\times N}\boldsymbol{\Gamma}_s\boldsymbol{H}_{N\times M}^H. \tag{2.91}$$

For example, say we estimate $\boldsymbol{X}$ by applying a filter $\boldsymbol{W}_{M\times M}^H$ on $\boldsymbol{Y}$,

$$\boldsymbol{W}_{M\times M}^H\boldsymbol{Y} = \widehat{\boldsymbol{X}} = \boldsymbol{W}_{M\times M}^H(\boldsymbol{X} + \boldsymbol{N}). \tag{2.92}$$

The power spectral density estimate at time-frame *r* and frequency-bin *k* is thus

$$\boldsymbol{\Gamma}_{\hat{x}} = \boldsymbol{W}_{M\times M}^H\big(\boldsymbol{\Gamma}_x + \boldsymbol{\Gamma}_\eta\big)\boldsymbol{W}_{M\times M}. \tag{2.93}$$

2.4.1 Noise reduction with Wiener filtering

The Wiener filter algorithm here is shown only to stress on the importance of estimating correctly the noise PSD matrix beforehand. Say we wish to find a matrix filter that will estimate $\boldsymbol{X}$ from the values of $\boldsymbol{Y}$ knowing that $\boldsymbol{\Gamma}_y$ is known.

We start by supposing that we know the noise PSD matrix. Or else, the MSE used in the following equation ends up with unknowns and can't be solved directly:

$$\boldsymbol{\mathcal{E}}[(\boldsymbol{W}_{M\times M}^H\boldsymbol{Y} - \boldsymbol{X})(\boldsymbol{W}_{M\times M}^H\boldsymbol{Y} - \boldsymbol{X})^H] = \boldsymbol{J}. \tag{2.94}$$

Taking the gradient with respect to the conjugate of $\boldsymbol{W}_{M\times M}$ we get,

$$\nabla_{\boldsymbol{W}_{M\times M}^*}\boldsymbol{J} = \boldsymbol{\mathcal{E}}[\boldsymbol{Y}(\boldsymbol{Y}^H\boldsymbol{W}_{M\times M} - \boldsymbol{X}^H)] = \boldsymbol{\Gamma}_y\boldsymbol{W}_{M\times M} - \boldsymbol{\Gamma}_{xy} = \boldsymbol{0},$$

$$\boldsymbol{\Gamma}_y\boldsymbol{W}_{M\times M} - \boldsymbol{\Gamma}_{xy} = \boldsymbol{0},$$

$$\rightarrow \boldsymbol{W}_{M\times M} = \boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_{xy},$$

and by independence of the noise $\boldsymbol{N}$ with $\boldsymbol{X}$,

$$\rightarrow \boldsymbol{W}_{M\times M} = \boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_x. \tag{2.95}$$

Since we suppose that the noise PSD matrix $\boldsymbol{\Gamma}_\eta$ is known, then we have no unknowns. Thus the preceding equation becomes:

$$\boldsymbol{W}_{M\times M} = \boldsymbol{\Gamma}_y^{-1}\left(\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_\eta\right). \quad \textbf{2.96}$$

Let us define the column vector $\boldsymbol{e}_n$ where the nth entry is 1 and all the other entries are 0's. We see in Eq.(2.96) that the Wiener filter depends on $\boldsymbol{\Gamma}_\eta$. We can use the vector $\boldsymbol{e}_n = \left[\delta[n-i]\right]_{i\in[0,M]}$ to select the nth column of $\boldsymbol{W}_{M\times M}$,

$$\boldsymbol{e}_n^T\boldsymbol{W}_{M\times M} = \boldsymbol{e}_n^T\boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_x = \boldsymbol{w}_n. \quad \textbf{2.97}$$

Each filter $\boldsymbol{w}_n$ will give the estimate of the nth entry of $\boldsymbol{X}$ i.e.

$$\boldsymbol{w}_n^T\boldsymbol{Y} = \hat{X}_n. \quad \textbf{2.98}$$

Note that there are many ways of estimating $\boldsymbol{X}$ depending on *a priori* information that we have about the variables, and the Wiener filter approach is only one of several methods. But discussing the different estimation algorithms for $\boldsymbol{X}$ is not the aim of the present work.

2.4.2 Performance measures

2.4.2.1 Input SNR

The overall input signal to noise ratio (iSNR) in dB is defined as:

$$iSNR_{dB} = 10\log_{10}\left[\frac{\sum_{n,i}x_i^2[n]}{\sum_{n,i}\eta_i^2[n]}\right], \quad \textbf{2.99}$$

where i is summed on all canals and n over all time samples.

2.4.2.2 Log-error

A mono-channel measure is the log-error [20, 21], where the average is over all frequency bins and time frames:

$$logErr_{dB} = \frac{1}{\mathcal{R}\mathcal{K}}\sum_{k,r}\left|10\log_{10}\left[\frac{\Gamma_\eta[r,k]}{\hat{\Gamma}_\eta[r,k]}\right]\right|. \quad \textbf{2.100}$$

and where the reference noise PSD $\Gamma_\eta[r,k]$ is directly measured when there is pure noise i.e., when we know the noise and read it directly from the sourceless $\Gamma_y[r,k]$.

Before averaging on the time and frequencies we have the time-frequency dependent log-error given by:

$$logErr_{dB}[r,k] = 10\log_{10}\left[\frac{\Gamma_\eta[r,k]}{\hat{\Gamma}_\eta[r,k]}\right]. \quad \textbf{2.101}$$

This measure is very well suited for mono-channel algorithms. However, for multichannel algorithms there is no unique reference Γ_η since it is almost certainly different for every channel. If we suppose that the noise field is homogeneous, it is intuitive to compare the estimated $\hat{\Gamma}_\eta$ with the noise psd of every channel. This measure is defined by the following equation:

$$logErr_{dB}[r,k] = \sum_i \left| 10\log_{10}\left[\frac{\Gamma_{\eta_i}[r,k]}{\hat{\Gamma}_{\eta}[r,k]}\right]\right|, \quad \textbf{2.102}$$

where the sum is taken over all channels. If the number of channels is odd then the minimum of the above equation is obtained whenever $\hat{\Gamma}_{\eta}$ is the median of the set of Γ_{η_i}'s. If the number of channels is even i.e., $M = 2n$, then the minimum is attained whenever $\hat{\Gamma}_{\eta}$ takes any value between the n[th] and the next largest value from the sorted set of Γ_{η_i}'s.In other words the minimum is in the median region. The minimum will thus not be unique except when the number of channels is odd since the median region is effectively a point.

One generalization of Eq.(2.100) to the multi-channel case is

$$logErr_{dB} = \frac{1}{M\mathcal{R}\mathcal{K}}\sum_{k,r}\sum_i \left| 10\log_{10}\left[\frac{\Gamma_{\eta_i}[r,k]}{\hat{\Gamma}_{\eta}[r,k]}\right]\right|, \quad \textbf{2.103}$$

where $\mathcal{R}$ is the total number of time frames and $\mathcal{K}$ is the number of frequency bins. The generalization measures on average how close our estimates are to the median region of the Γ_{η_i}'s. If the noise field is inhomogeneous, we will estimate the noise PSDs out of the formula (see Eq.(2.25))

$$\hat{\Gamma}_{\eta_i} = \hat{\Gamma}_{v}\Psi_{\eta_i}, \quad \textbf{2.104}$$

and the performance measures generalize for the inhomogeneous case to:

$$logErr_{dB} = \frac{1}{M\mathcal{R}\mathcal{K}}\sum_{k,r}\sum_i \left| 10\log_{10}\left[\frac{\Gamma_{\eta_i}[r,k]}{\hat{\Gamma}_{\eta_i}[r,k]}\right]\right|. \quad \textbf{2.105}$$

One important problem with all the measures above is that they don't take in account errors on the correlation or coherence matrix that we use to compute $\hat{\Gamma}_{\eta}$ and to ultimately estimate $\mathbf{\Gamma}_{\eta}$. We have assumed since the beginning that we somehow know the coherences or correlation matrix of the noise field. But those matrices are obviously prone to error. So it would be better to measure how close is our estimated noise PSD matrix to the actual noise PSD matrix or, how close is $\hat{\mathbf{\Gamma}}_{\eta}$ to the actual $\mathbf{\Gamma}_{\eta}$.

2.4.2.3 Square error of the noise power matrix

One possible measure is the average on all time and frequencies of the Frobenius norm of the difference of the matrices $\hat{\mathbf{\Gamma}}_{\eta}$ and $\mathbf{\Gamma}_{\eta}$:

$$\begin{aligned} MSE\left(\hat{\mathbf{\Gamma}}_{\eta}, \mathbf{\Gamma}_{\eta}\right) &= \frac{1}{M^2\mathcal{R}\mathcal{K}}\sum_{k,r,i,j}\left|\Gamma_{\eta_i\eta_j}[r,k] - \hat{\Gamma}_{\eta_i\eta_j}[r,k]\right|^2 \\ &= \frac{1}{M^2\mathcal{R}\mathcal{K}}\sum_{k,r,i,j} tr\left(\left(\boldsymbol{\Gamma}_{\eta}[r,k] - \hat{\boldsymbol{\Gamma}}_{\eta}[r,k]\right)^2\right) \end{aligned} \quad \textbf{2.106}$$

Although it is a good measure since it takes into consideration off diagonal terms, it doesn't generalize Eq.(2.100). There are multiple possibilities to generalize the desired equation.

2.4.2.4 Multichannel log-error(★):

One possibility to generalize Eq.(2.100) is to use the following equation:

$$d(\hat{\mathbf{\Gamma}}_\eta, \mathbf{\Gamma}_\eta) = \frac{1}{M\mathcal{R}\mathcal{K}} \sum_{k,r} tr\left(\left|\log(\boldsymbol{\Gamma}_\eta[r,k]) - \log(\hat{\boldsymbol{\Gamma}}_\eta[r,k])\right|\right). \quad \textbf{2.107}$$

The summand $tr\left(\left|\log(\boldsymbol{\Gamma}_\eta[r,k]) - \log(\hat{\boldsymbol{\Gamma}}_\eta[r,k])\right|\right)$ is understood as being the sum of the absolute value of the eigenvalues of $\left|\log(\boldsymbol{\Gamma}_\eta[r,k]) - \log(\hat{\boldsymbol{\Gamma}}_\eta[r,k])\right|$, and here the logarithms are legal operations on the matrices since they are positive definite [13]. In dB the measure can be rewritten as

$$d_{dB}(\hat{\mathbf{\Gamma}}_\eta, \mathbf{\Gamma}_\eta) = \frac{1}{M\mathcal{R}\mathcal{K}} \sum_{k,r} tr\left(10\left|\log_{10}(\boldsymbol{\Gamma}_\eta[r,k]) - \log_{10}(\hat{\boldsymbol{\Gamma}}_\eta[r,k])\right|\right). \quad \textbf{2.108}$$

Note that Eq.(2.103, 2.105) can be derived from the last equation if we completely ignore the contribution of the off-diagonal terms i.e.,

$$logErr_{dB} = d_{dB}(\hat{\mathbf{\Gamma}}_\eta \circ \boldsymbol{I}, \mathbf{\Gamma}_\eta \circ \boldsymbol{I})$$
$$= \frac{1}{M\mathcal{R}\mathcal{K}} \sum_{k,r} tr\left(10\left|\log_{10}(\boldsymbol{\Gamma}_\eta[r,k] \circ \boldsymbol{I}) - \log_{10}(\hat{\boldsymbol{\Gamma}}_\eta[r,k] \circ \boldsymbol{I})\right|\right). \quad \textbf{2.109}$$

2.4.2.5 Squared Log-error

Keeping in mind that we wish to estimate the noise matrix, we have another measure called the mean squared log-error:

$$MSlogE(\hat{\mathbf{\Gamma}}_\eta, \mathbf{\Gamma}_\eta) = \frac{1}{M^2\mathcal{R}\mathcal{K}} \sum_{k,r} tr\left(\left(\log(\boldsymbol{\Gamma}_\eta[r,k]) - \log(\hat{\boldsymbol{\Gamma}}_\eta[r,k])\right)^2\right). \quad \textbf{2.110}$$

The squared log-error is related to the squared error measure defined earlier in Eq.(2.106) with,

$$MSlogE(\hat{\mathbf{\Gamma}}_\eta, \mathbf{\Gamma}_\eta) = MSE(\log(\hat{\boldsymbol{\Gamma}}_\eta), \log(\boldsymbol{\Gamma}_\eta)). \quad \textbf{2.111}$$

2.5 Algebraic solution for the homogeneous case

First off, we consider the case where there are less sources than microphones or $N < M$. Then we suppose that all noise PSDs are the same for each microphone, i.e., the noise field is homogeneous with $\Gamma_{\eta_i} = \Gamma_{\eta_j} = \Gamma_\eta$. We thus have the following system,

$$\mathbf{\Gamma}_y = \mathbf{\Gamma}_x + \Gamma_\eta \mathbf{\Phi}. \quad \textbf{2.112}$$

The solution to this problem is the minimal generalized eigenvalue that solves the equation:

$$\chi_{\boldsymbol{\Gamma}_y,\boldsymbol{\Phi}}(\Gamma_\eta) = \det(\Gamma_\eta \boldsymbol{\Phi} - \boldsymbol{\Gamma}_y) = 0, \tag{2.113}$$

where $\chi_{\boldsymbol{A},\boldsymbol{B}}(\lambda)$ is called the characteristic polynomial of the pencil $\lambda\boldsymbol{B} - \boldsymbol{A}$ [13]. It is interesting to note that the minimal eigenvalue solution is reminiscent of the minimal statistics algorithm. Two proofs will be presented below and each is important as this leads to two types of algorithms.

First proof

The matrix $\boldsymbol{\Gamma}_x$ is rank defective i.e., $rank(\boldsymbol{\Gamma}_x) = M - d$. This implies that the defect number $def(\boldsymbol{\Gamma}_x) = d > 0$. The defect number of a matrix is defined in [13] as the dimension of the null space of a given matrix $\boldsymbol{A}$ or, $def(\boldsymbol{A}) = \dim(\mathcal{N}(\boldsymbol{A}))$.

Let the quantity $\hat{\Gamma}_\eta$ be the solution to the problem then we have

$$\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_x + \hat{\Gamma}_\eta \boldsymbol{\Phi}. \tag{2.114}$$

If we subtract from the matrix equation 2.114 the term $\Gamma_\eta \boldsymbol{\Phi}$, where Γ_η is a variable to be determined, we have

$$\boldsymbol{\Gamma}_y - \Gamma_\eta \boldsymbol{\Phi} = \boldsymbol{\Gamma}_x + (\hat{\Gamma}_\eta - \Gamma_\eta)\boldsymbol{\Phi}. \tag{2.115}$$

If $\boldsymbol{\Phi}$ is symmetric, then the coherence matrix is factorable under the form $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2}$. Then, the above matrix can be brought to the form

$$\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_y\boldsymbol{\Phi}^{-1/2} - \Gamma_\eta \mathbf{I} = \boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2} + (\hat{\Gamma}_\eta - \Gamma_\eta)\mathbf{I}. \tag{2.116}$$

The above process is equivalent to whitening the noise field. If we take the determinant on both sides of the equation, we are brought to

$$\begin{aligned} &\det(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_y\boldsymbol{\Phi}^{-1/2} - \Gamma_\eta \mathbf{I}) \\ &\qquad = \det(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2} - (\Gamma_\eta - \hat{\Gamma}_\eta)\boldsymbol{I}). \end{aligned} \tag{2.117}$$

Now to find a solution for $\det(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_y\boldsymbol{\Phi}^{-1/2} - \Gamma_\eta \mathbf{I}) = 0$, the above vanishing polynomial can be rewritten as

$$\prod_{i=1}^{M} (\lambda_i(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2}) - \Gamma_\eta + \hat{\Gamma}_\eta) = 0, \tag{2.118}$$

where the eigenvalues of $\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2}$ are denoted $\lambda_i(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2}) = \lambda_i = \lambda_i(\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}_x)$ for concision purposes. The rank defectiveness assumption implies that

$$\text{def}(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-1/2}) = \text{def}(\boldsymbol{\Gamma}_x) = d = M - N. \tag{2.119}$$

Also, let the eigenvalues λ_i be ordered in a decreasing order way i.e.,

$$\lambda_1 \geq \cdots \geq \lambda_{M-d} > \lambda_{M-d+1} = \cdots = \lambda_M = 0. \tag{2.120}$$

Now, it can be seen that the above polynomial can be factored in the following way

$$\left(\hat{\Gamma}_\eta - \Gamma_\eta\right)^d \prod_{i=1}^{M-d}\left(\lambda_i - \Gamma_\eta + \hat{\Gamma}_\eta\right) = 0. \quad \textbf{2.121}$$

For Γ_η to be the solution that is $\Gamma_\eta = \hat{\Gamma}_\eta$, by inspection we see that the only solution to the above product is the one that has the minimal value. This proof is equivalent to the reasoning in [3] that showed why the negative root solution (minimal solution) was chosen for $M = 2$, $N = 1$. It is important to notice that in Eq.(2.121) the minimum eigenvalue has multiplicity d. In practice this is highly unlikely, but it would not alter the solution. This proof leads us to compute the minimum root of the polynomial $\det\left(\mathbf{\Gamma}_y - \Gamma_\eta \mathbf{\Phi}\right)$.

Second proof

Now say we wish to find a simple expression for Γ_η. To do this we suppose that there exists a vector $\boldsymbol{v} \in \mathbb{C}^M$ that corresponds to a filter that when applied to $\boldsymbol{Y}$ leads to:

$$\boldsymbol{v}^H \boldsymbol{Y} = \boldsymbol{v}^H (\boldsymbol{X} + \boldsymbol{N}). \quad \textbf{2.122}$$

The variance of the above equation gives:

$$\boldsymbol{v}^H \mathbf{\Gamma}_y \boldsymbol{v} = \boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v} + \Gamma_\eta \boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}. \quad \textbf{2.123}$$

Notice that by definition of the case considered, $\mathbf{\Gamma}_x$ is positive semi-definite whereas $\mathbf{\Gamma}_y$ and $\mathbf{\Phi}$ are positive definite. So by definition $\boldsymbol{v}^H \mathbf{\Gamma}_y \boldsymbol{v} > 0, \boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v} > 0, \boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v} \geq 0$. Dividing the preceding equation by $\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}$ is thus possible since the quadratic form $\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v} \neq 0$ and we get:

$$\frac{\boldsymbol{v}^H \mathbf{\Gamma}_y \boldsymbol{v}}{\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}} = \frac{\boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}}{\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}} + \Gamma_\eta. \quad \textbf{2.124}$$

Now we wish to find a solution for Γ_η that minimises the energy contribution of $\mathbf{\Gamma}_x$ to Γ_η because the signal and the noise are assumed to be statistically independent. So we get a hint that we need to find a vector $\boldsymbol{v}$ that minimizes $\frac{\boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}}{\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}}$ or,

$$\arg\min_{\boldsymbol{v}} \frac{\boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}}{\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}}. \quad \textbf{2.125}$$

Now since $\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}$ is positive definite we know that only $\boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}$ will be the factor that can give a zero since it is positive semi-definite by hypothesis, that is,

$$\arg\min_{\boldsymbol{v}} \frac{\boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}}{\boldsymbol{v}^H \mathbf{\Phi} \boldsymbol{v}} = \arg\min_{\boldsymbol{v}} \boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v}. \quad \textbf{2.126}$$

Since $\mathbf{\Gamma}_x$ decomposes into $\boldsymbol{H}_{M\times N} \mathbf{\Gamma}_s \boldsymbol{H}_{N\times M}^H$, we have

$$\arg\min_{\boldsymbol{v}} \boldsymbol{v}^H \mathbf{\Gamma}_x \boldsymbol{v} = \arg\min_{\boldsymbol{v}} \boldsymbol{v}^H \boldsymbol{H}_{M\times N} \mathbf{\Gamma}_s \boldsymbol{H}_{N\times M}^H \boldsymbol{v}. \quad \textbf{2.127}$$

We now suppose that a minimum exists and it is 0, this can only be true if $\boldsymbol{v}$ is in the null space of $\boldsymbol{H}_{N\times M}^{H}$, or equivalently orthogonal to the range of $\boldsymbol{H}_{M\times N}$. In other terms,

$$\min \boldsymbol{v}^{H}\boldsymbol{H}_{M\times N}\boldsymbol{\Gamma}_{s}\boldsymbol{H}_{N\times M}^{H}\boldsymbol{v}=0 \leftrightarrow \boldsymbol{v}\in \mathcal{R}^{\perp}(\boldsymbol{H}_{M\times N})=\mathcal{N}\left(\boldsymbol{H}_{N\times M}^{\boldsymbol{H}}\right), \tag{2.128}$$

where $\mathcal{N}(\boldsymbol{A})$ denotes the null space of $\boldsymbol{A}$ and $\mathcal{R}^{\perp}(\boldsymbol{A})$ denotes the space orthogonal to the span of the columns of $\boldsymbol{A}$**.** We then get the following equation to find the possible minimizing vectors,

$$\min \frac{\boldsymbol{v}^{H}\boldsymbol{\Gamma}_{y}\boldsymbol{v}}{\boldsymbol{v}^{H}\boldsymbol{\Phi}\boldsymbol{v}}=\min \Gamma_{\eta}. \tag{2.129}$$

We need to set the gradient $\nabla_{v^{*}}\,\Gamma_{\eta}=\mathbf{0}$ using Wirtinger's calculus[22] to find the optimal points:

$$\begin{aligned}\nabla_{v^{*}}\left(\Gamma_{\eta}\boldsymbol{v}^{H}\boldsymbol{\Phi}\boldsymbol{v}\right)&=\nabla_{v^{*}}\,\boldsymbol{v}^{H}\boldsymbol{\Gamma}_{y}\boldsymbol{v},\\ \rightarrow \nabla_{v^{*}}\,\Gamma_{\eta}+\nabla_{v^{*}}\boldsymbol{v}^{H}\boldsymbol{\Phi}\boldsymbol{v}&=\nabla_{v^{*}}\,\boldsymbol{v}^{H}\boldsymbol{\Gamma}_{y}\boldsymbol{v},\\ \rightarrow \nabla_{v^{*}}\,\Gamma_{\eta}+\Gamma_{\eta}\boldsymbol{\Phi}\boldsymbol{v}&=\boldsymbol{\Gamma}_{y}\boldsymbol{v},\end{aligned} \tag{2.130}$$

Inserting the optimality condition $\nabla_{v^{*}}\,\Gamma_{\eta}=\mathbf{0}$, we get:

$$\Gamma_{\eta}\boldsymbol{\Phi}\boldsymbol{v}=\ \boldsymbol{\Gamma}_{y}\boldsymbol{v}. \tag{2.131}$$

This is a classical generalized eigenvalue problem where we want to find the minimal value of Γ_{η} (a generalized Rayleigh quotient) that solves the following determinant equation,

$$\det\left(\Gamma_{\eta}\boldsymbol{\Phi}-\boldsymbol{\Gamma}_{y}\right)=\ 0. \tag{2.132}$$

The relation of Γ_{η} as being a generalized Rayleigh quotient leads us to an important iterative algorithm to compute Γ_{η}. Using the notation $\lambda_{min}\left(\boldsymbol{\Gamma}_{y},\boldsymbol{\Phi}\right)$ to design the minimal generalized eigenvalue the estimate of Γ_{η} is

$$\lambda_{min}\left(\boldsymbol{\Gamma}_{y},\boldsymbol{\Phi}\right)=\hat{\Gamma}_{\eta}, \tag{2.133}$$

and the noise PSD matrix estimate is thus

$$\hat{\Gamma}_{\eta}\boldsymbol{\Phi}=\lambda_{min}\left(\boldsymbol{\Gamma}_{y},\boldsymbol{\Phi}\right)\boldsymbol{\Phi}=\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}. \tag{2.134}$$

In this section, we note that the minimal eigenvalue has multiplicity $M-N$.

2.6 Generalized eigenvalue sensitivity issues

In [23] it is mentioned that the chordal metric $cord(a,b)$ is an appropriate measure of the generalized eigenvalue perturbation:

$$cord(a,b)=\frac{|a-b|}{\sqrt{1+a^{2}}\sqrt{1+b^{2}}}. \tag{2.135}$$

In [24] it is shown that if λ is an eigenvalue of $\boldsymbol{A}-\lambda\boldsymbol{B}$ and λ_ϵ is the eigenvalue of the perturbed pencil $\boldsymbol{A}+\Delta\boldsymbol{A}-\lambda(\boldsymbol{B}+\Delta\boldsymbol{B})$ with $\|\Delta\boldsymbol{A}\|_2 \approx \|\Delta\boldsymbol{B}\|_2 \approx \epsilon$, then

$$cord(\lambda,\lambda_\epsilon) \leq \frac{\epsilon}{\sqrt{(\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x})^2+(\boldsymbol{y}^H\boldsymbol{B}\boldsymbol{x})^2}} + O(\epsilon^2) \tag{2.136}$$

where x and y^H are the right and left eigenvectors[5] of $\boldsymbol{A}-\lambda\boldsymbol{B}$ respectively with unit 2-norms. The ill-conditioned eigenvalues are those that have a large denominator $\sqrt{(\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x})^2+(\boldsymbol{y}^H\boldsymbol{B}\boldsymbol{x})^2}$ [23]. Using the notation of this section, the above metric becomes:

$$cord\left(\Gamma_\eta,\Gamma_{\eta_\epsilon}\right) \leq \frac{\epsilon}{\sqrt{\left(\boldsymbol{v}^H\boldsymbol{\Gamma}_y\boldsymbol{v}\right)^2+(\boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{v})^2}} + O(\epsilon^2) \tag{2.137}$$

where $\left\|\Delta\boldsymbol{\Gamma}_y\right\|_2 \approx \|\Delta\boldsymbol{\Phi}\|_2 \approx \epsilon$ is small. By Eq.(2.128), v is in $\mathcal{N}(\boldsymbol{\Gamma}_x)$ hence the above equation reduces to:

$$\begin{aligned} cord\left(\Gamma_\eta,\Gamma_{\eta_\epsilon}\right) &\leq \frac{\epsilon}{\sqrt{\Gamma_\eta^2(v^H\boldsymbol{\Phi}\boldsymbol{v})^2+(\boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{v})^2}} + O(\epsilon^2) \\ &= \frac{\epsilon}{\boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{v}\sqrt{\Gamma_\eta^2+1}} + O(\epsilon^2) \end{aligned} \tag{2.138}$$

This means that if the measurement error on Γ_y or Φ are comparable and small and if $\boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{v}$ is large (compared to the error), then the perturbed eigenvalue is only lightly affected by the perturbations. However in practice, we don't know if $\left\|\Delta\boldsymbol{\Gamma}_y\right\|_2 \approx \|\Delta\boldsymbol{\Phi}\|_2 \approx \epsilon$ is true nor do we know the order of magnitude of ϵ.

2.7 Sample matrix estimate of $\boldsymbol{\Gamma}_y$

The matrix $\boldsymbol{\Gamma}_y$ is estimated by either using recursive or non-recursive methods by using respectively

$$\hat{\boldsymbol{\Gamma}}_y[r,k] = \alpha\hat{\boldsymbol{\Gamma}}_y[r-1,k] + (1-\alpha)\boldsymbol{Y}[r,k]\boldsymbol{Y}^H[r,k], \tag{2.139}$$

and

$$\hat{\boldsymbol{\Gamma}}_y[r,k] = \frac{1}{L}\sum_{i=0}^{L-1}\boldsymbol{Y}[r-i,k]\boldsymbol{Y}^H[r-i,k]. \tag{2.140}$$

[5] Left and right eigenvectors are complex conjugates when the matrices *A* and *B* are Hermitian.

2.8 Signal subspace dimensionality estimation

Because of the fact that we use samples to estimate $\boldsymbol{\Gamma}_y$, the algebraic solution obtained by taking the minimal eigenvalue might not be the best one. Assuming Gaussianity[6] of the observations, Anderson [26] has shown that if we use L snapshots in Eq.(2.140) to estimate $\boldsymbol{\Gamma}_y$ then, the $M-N$ smallest eigenvalues cluster around the true noise variance. Using our notation that is:

$$\hat{\lambda}_m - \Gamma_\eta = O\left(L^{-\frac{1}{2}}\right); m \in [N+1, M]. \quad \textbf{2.141}$$

Anderson [26] also showed that a sufficient statistic is:

$$L_N(\widehat{N}) = L(M-\widehat{N}) \ln\left(\frac{\frac{1}{M-\widehat{N}}\sum_{m=\widehat{N}+1}^{M}\hat{\lambda}_m}{\sqrt[M-\widehat{N}]{\prod_{m=\widehat{N}+1}^{M}\hat{\lambda}_m}}\right). \quad \textbf{2.142}$$

The maximum likelihood (ML) of Γ_η is given by [22]:

$$\hat{\Gamma}_\eta = \frac{1}{M-\widehat{N}}\sum_{m=\widehat{N}+1}^{M}\hat{\lambda}_m. \quad \textbf{2.143}$$

We can observe that when the $M-\widehat{N}$ eigenvalues are equal, $L_N(\widehat{N}) = 0$. Gupta [27] has shown that $2L_N(\widehat{N})$ corresponds to a chi-squared random variable $\chi^2((M-N)^2-1)$. Bartlett [28] and Lawley [29] developed the sequential hypothesis test (SHT) where the values of $L_N(N)$ are sequentially tested with their assumed model to determine a confidence bound estimate. For example, using a confidence interval of 99% (chosen arbitrarily) and computing the corresponding threshold $\gamma_{99}^{(N)}$, we can then proceed through each value of N to find the estimate $\widehat{N}_{SHT}$ with,

$$\widehat{N}_{SHT} = \arg\min_{\widetilde{N}} \widetilde{N} \in \left[L_N(\widehat{N}) \le \gamma_{99}^{(\widehat{N})}\left(\chi^2((M-N)^2-1)\right)\right]; \widehat{N} \in [0, M-1]. \quad \textbf{2.144}$$

In [30], it is shown that for a low sample support ($L/M \le 5$) $L_N(\widehat{N})$ will not accurately have the prescribed distribution and hence, we can no longer rely on the confidence interval to estimate the number of signals. This is a very important issue since short-time correlation matrix estimates will necessarily use a small L (in Eq.(2.140)) or a small α (in Eq.(2.139)). This is one of the issues that motivated the development of the Akaike Information Criterion (AIC) and the minimum description length (MDL) information-theoretic detection schemes described respectively in [31] and [32, 33]. In [22] it is mentioned that the AIC has a higher probability of correct decision for low values of L compared to the MDL estimate. The AIC test is

$$AIC(\widehat{N}) \triangleq L_N(\widehat{N}) + \widehat{N}(2M-\widehat{N}), \quad \textbf{2.145}$$

[6] Since sources are not Gaussian in general [25], it would be preferable to use a super-Gaussian pdf model for the principal component analysis used in deriving a sufficient statistic for the signal subspace dimension.

and the estimated number of signals is:

$$\widehat{N}_{AIC} = \arg\min_{\widetilde{N}} AIC(\widetilde{N}). \quad \textbf{2.146}$$

If $\alpha = 0.7$ is used, this will act as if we used a small number of samples to estimate $\mathbf{\Gamma}_y$ and therefore it would be advisable to use the AIC to estimate N. On the contrary, if $\alpha = 0.9$ then it would correspond to a larger equivalent number of samples L and it would be better to use the MDL to estimate N since the AIC is biased for a large support L/M[30]. The MDL test is

$$MDL(\widehat{N}) \triangleq L_N(\widehat{N}) + \frac{\ln(L)}{2}\left(\widehat{N}(2M - \widehat{N}) + 1\right), \quad \textbf{2.147}$$

and the estimated number of signals is:

$$\widehat{N}_{MDL} = \arg\min_{\widetilde{N}} MDL(\widetilde{N}). \quad \textbf{2.148}$$

2.9 Eigenvalue closed form expressions

In this section eigenvalue root formulas for $M = (2,3,4)$ will be shown. Since we use the notation that eigenvalues are ordered in decreasing fashion, the eigenvalues Γ_η will be ordered as such with a superscript, i.e., $\Gamma_\eta^{(1)} > \Gamma_\eta^{(2)} > \cdots > \Gamma_\eta^{(M)}$, not to be confused with the notation for noise PSD matrix estimations versions $\widehat{\mathbf{\Gamma}}_\eta^{(1)}$, $\widehat{\mathbf{\Gamma}}_\eta^{(2)}$,and $\widehat{\mathbf{\Gamma}}_\eta^{(3)}$ used later in the thesis.

Two microphones case:

If we compute $\det\left(\Gamma_\eta \boldsymbol{\Phi} - \boldsymbol{\Gamma}_y\right)$ we get:

$$\det\left(\Gamma_\eta \boldsymbol{\Phi} - \boldsymbol{\Gamma}_y\right) = \Gamma_\eta^2 \det(\boldsymbol{\Phi}) - \Gamma_\eta \left(tr(\boldsymbol{\Phi})tr\left(\boldsymbol{\Gamma}_y\right) - tr\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y\right)\right) + \det\left(\boldsymbol{\Gamma}_y\right),$$

or after simplification, **2.149**

$$\det\left(\Gamma_\eta \boldsymbol{\Phi} - \boldsymbol{\Gamma}_y\right) = \Gamma_\eta^2 \left(1 - |\Phi|^2\right) - \Gamma_\eta \left(\Gamma_{y_1} + \Gamma_{y_2} - 2\Phi Re\left(\Gamma_{y_{12}}\right)\right) + \Gamma_{y_1}\Gamma_{y_2} - \left|\Gamma_{y_{12}}\right|^2.$$

Setting the polynomial to zero and solving for the roots, we obtain:

If we use,

$$b = 2Re\left(\Phi^* \Gamma_{y_{12}}\right) - \Gamma_{y_1} - \Gamma_{y_2},$$
$$\Delta = \left(\Gamma_{y_1} + \Gamma_{y_2} - 2Re\left(\Phi^* \Gamma_{y_{12}}\right)\right)^2 - 4\left(\Gamma_{y_1}\Gamma_{y_2} - \left|\Gamma_{y_{12}}\right|^2\right)\left(1 - |\Phi|^2\right), \quad \textbf{2.150}$$

the eigenvalues become:

$$\Gamma_\eta^{(1)} = \frac{\left(\Gamma_{y_1} + \Gamma_{y_2} - 2Re\left(\Phi^* \Gamma_{y_{12}}\right) + \sqrt{\Delta}\right)}{2\left(1 - |\Phi|^2\right)}, \quad \textbf{2.151}$$

$$\Gamma_\eta^{(2)} = \frac{\left(\Gamma_{y_1} + \Gamma_{y_2} - 2Re\left(\Phi^* \Gamma_{y_{12}}\right) - \sqrt{\Delta}\right)}{2\left(1 - |\Phi|^2\right)}. \quad \textbf{2.152}$$

Three microphones case:

When $M = 3$, the characteristic polynomial becomes:

$$\begin{aligned}
\det\left(\Gamma_\eta \boldsymbol{\Phi} - \boldsymbol{\Gamma}_y\right) &= a\Gamma_\eta^3 + b\Gamma_\eta^2 + c\Gamma_\eta + d, \\
a &= \det(\boldsymbol{\Phi}), \\
b &= -\left(tr\left(\boldsymbol{\Gamma}_y\boldsymbol{\Phi}^2\right) - tr\left(\boldsymbol{\Gamma}_y\boldsymbol{\Phi}\right)tr(\boldsymbol{\Phi}) + \frac{1}{2}\left(tr\left(\boldsymbol{\Gamma}_y\right)tr^2(\boldsymbol{\Phi}) - tr\left(\boldsymbol{\Gamma}_y\right)tr(\boldsymbol{\Phi}^2)\right)\right) \\
&= -tr\left(\boldsymbol{\Phi}^A\boldsymbol{\Gamma}_y\right), \\
c &= tr\left(\boldsymbol{\Gamma}_y{}^2\boldsymbol{\Phi}\right) - tr\left(\boldsymbol{\Gamma}_y\boldsymbol{\Phi}\right)tr\left(\boldsymbol{\Gamma}_y\right) + \frac{1}{2}\left(tr^2\left(\boldsymbol{\Gamma}_y\right)tr(\boldsymbol{\Phi}) - tr\left(\boldsymbol{\Gamma}_y{}^2\right)tr(\boldsymbol{\Phi})\right) = tr\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y^A\right), \\
d &= -\det\left(\boldsymbol{\Gamma}_y\right),
\end{aligned} \tag{2.153}$$

where,

$$\boldsymbol{B}^A = \boldsymbol{B}^2 - tr(\boldsymbol{B})\boldsymbol{B} + \frac{1}{2}\left(tr^2(\boldsymbol{B}) - tr(\boldsymbol{B}^2)\right)\boldsymbol{I}.$$

The above formulas for the coefficients were obtained using the Cayley-Hamilton theorem (CHT) with $\Gamma_\eta\boldsymbol{\Phi} - \boldsymbol{\Gamma}_y$. We present the eigenvalues in the case of $M = 3$. Note that the notation in terms of traces and adjoints is a compact one, but care must be taken when computing the coefficients as the notation might lead to unnecessary operations since some simplifications are possible, for example $tr(\boldsymbol{\Phi}) = 3$. If we ignore the fact that the diagonals of $\boldsymbol{\Phi}$ are unity, computing each coefficient should have a complexity of $O(M^3)$ and since there are $M + 1$ coefficients, the total complexity to compute the coefficients is of $O(M^4)$. Even with simplifications done, the complexity is quite high. The derivation of the cubic minimal solution is done in appendix D.

$$\Gamma_\eta^{(1)} = -\frac{b}{3a} + \frac{2}{3a}\sqrt{b^2 - 3ac}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3^3qa^3}{2\sqrt{(b^2 - 3ac)^3}}\right)\right), \tag{2.154}$$

$$\Gamma_\eta^{(2)} = -\frac{b}{3a} + \frac{2}{3a}\sqrt{b^2 - 3ac}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3^3qa^3}{2\sqrt{(b^2 - 3ac)^3}}\right) + \frac{4\pi}{3}\right), \tag{2.155}$$

$$\Gamma_\eta^{(3)} = -\frac{b}{3a} + \frac{2}{3a}\sqrt{b^2 - 3ac}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3^3qa^3}{2\sqrt{(b^2 - 3ac)^3}}\right) + \frac{2\pi}{3}\right), \tag{2.156}$$

where

$$q = \frac{1}{3^3a^3}(2b^3 - 9abc + 27a^2d). \tag{2.157}$$

Four microphones case:

The complexity problem becomes worse for $M = 4$ as we shall see. For the generalized characteristic polynomial we have

$$\begin{aligned}
\det\left(\Gamma_\eta\boldsymbol{\Phi} - \boldsymbol{\Gamma}_y\right) &= a\Gamma_\eta^4 + b\Gamma_\eta^3 + c\Gamma_\eta^2 + d\Gamma_\eta + e, \\
a &= \det(\boldsymbol{\Phi}), \\
b &= -tr\left(\boldsymbol{\Phi}^A\boldsymbol{\Gamma}_y\right),
\end{aligned} \tag{2.158}$$

$$c = \frac{1}{4}\Big[tr^2(\boldsymbol{\Phi})tr^2(\boldsymbol{\Gamma}_y) - \Big(tr^2(\boldsymbol{\Phi})tr(\boldsymbol{\Gamma}_y^2) + 4tr(\boldsymbol{\Phi})tr(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y)tr(\boldsymbol{\Gamma}_y) + tr^2(\boldsymbol{\Gamma}_y)tr(\boldsymbol{\Phi}^2)\Big) - \Big(4tr(\boldsymbol{\Phi}^2\boldsymbol{\Gamma}_y^2) + 2tr(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y\boldsymbol{\Phi}\boldsymbol{\Gamma}_y)\Big) + \Big(2tr^2(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y) + tr(\boldsymbol{\Gamma}_y^2)tr(\boldsymbol{\Phi}^2)\Big) + 4\Big(tr(\boldsymbol{\Phi})tr(\boldsymbol{\Gamma}_y^2\boldsymbol{\Phi}) + tr(\boldsymbol{\Gamma}_y)tr(\boldsymbol{\Phi}^2\boldsymbol{\Gamma}_y)\Big)\Big],$$

$$d = -tr(\boldsymbol{\Phi}\boldsymbol{\Gamma}_y^A),$$

$$e = \det(\boldsymbol{\Gamma}_y),$$

where the adjoint matrix formula $\boldsymbol{B}^A$ in terms of the matrix $\boldsymbol{B}$ and its traces can be computed using the CHT for $\boldsymbol{B}$, multiplying both sides of the equation by $\boldsymbol{B}^A$, simplifying and then solving for $\boldsymbol{B}^A$. As can be seen, simplifying the above coefficients becomes difficult and may require the use of symbolic algebra handling tools. But even after full simplification, the results will have to be coded painstakingly since the expressions will be long and hence are prone to human implementation error. An alternative coefficient formula[7] is given in appendix E. Stated without proof is the formula for the roots. The solution was chosen from the four solutions shown in [34]. The amplitude order of the roots can be checked by numerical evaluation of the 4 formulas whenever all roots are real.

$$\Gamma_\eta^{(1)} = -\frac{b}{4a} + \frac{1}{2}R + \frac{1}{2}D, \quad \textbf{2.159}$$

$$\Gamma_\eta^{(2)} = -\frac{b}{4a} + \frac{1}{2}R - \frac{1}{2}D, \quad \textbf{2.160}$$

$$\Gamma_\eta^{(3)} = -\frac{b}{4a} - \frac{1}{2}R + \frac{1}{2}E, \quad \textbf{2.161}$$

$$\Gamma_\eta^{(4)} = -\frac{b}{4a} - \frac{1}{2}R - \frac{1}{2}E, \quad \textbf{2.162}$$

with,

$$R = \sqrt{-\frac{2}{3}p + \frac{2}{3a}\sqrt{\Delta_0}\cos(\psi)},$$

$$D = \sqrt{-R^2 - 2p + \frac{2q}{R}}, \quad \textbf{2.163}$$

$$E = \sqrt{-R^2 - 2p - \frac{2q}{R}},$$

$$p = \frac{8ac - 3b^2}{8a^2},$$

[7] This formula is based on the Leibnitz expression for determinants to compute the coefficients. The expression is easily implemented compared to the expressions using traces. Although the Leibnitz formula is inefficient for large values of M(complexity of $O(M!)$), it can be worthwhile using especially if one needs to compute eigenvalues on multiple time frames and frequency bins with an interpreted language such as MATLAB since the Leibnitz and closed form expressions for the eigenvalues are easily vectorized.

$$q = \frac{b^3 - 4abc + 8a^2 d}{8a^3},$$

$$\psi = \frac{1}{3}\text{acos}\left(\frac{\Delta_1}{2\sqrt{\Delta_0^3}}\right),$$

$$\Delta_0 = c^2 - 3bd + 12ae,$$

$$\Delta_1 = 2c^3 - 9bcd + 27b^2 e + 27ad^2 - 72ace.$$

From Eqs.(2.163) and (2.158) we can see that computing the solution will have a great complexity, this is largely because of the computation of the coefficients. For example the term $tr\left(\boldsymbol{\Phi\Gamma}_y\boldsymbol{\Phi\Gamma}_y\right)$ in the coefficient c requires $2M^3 + M^2 = 144$ complex multiplications and $3M^3 - 4M^2 + M = 132$ complex additions. Computing all the coefficients has a complexity of $O(M^4)$, this is very ineffective compared to the $O(M^3)$ required to compute a generalized eigenvalue [23].

Discussion of the closed form solutions:

For $M = 2$ there isn't too much to worry about computational complexity issues compared to the cases for $M = (3,4)$. Additionaly, the large amount of operations required to compute the coefficients for $M = (3,4)$ introduce errors in the coefficients since we use finite precision floating point arithmetic. If small perturbations on the coefficients lead to small perturbations on the estimate of the eigenvalue Γ_{η_i} and if processing time is not an issue, then we can use the above closed form expressions.

Assuming that we know the coefficients, if we were to use a root solver as in the MATLAB package the problem would become more sensitive to numerical errors since it computes the roots out of the eigenvalues of the companion matrix associated with the coefficients using iterative methods [35]. Having an analytical solution is numerically more reliable than this approach. But for a polynomial of degree 5 or more, it is not possible to use radicals or other elementary functions to express the roots in a closed form. This is a classical result in Galois theory. Hence we must use iterative algorithms to find the roots in such cases. The best bet would be to use iterative methods to compute the smallest eigenvalue using directly the matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}_y$ instead of the coefficients. For large polynomial degrees, polynomial root finding can become sensitive to numerical errors on the coefficients if the polynomial equation is ill-conditioned [35], see for example Wilkinson's polynomial. Because the error potentially accumulates using the analytical solution for $M = (3,4)$, the coefficient exact expressions are complicated to simplify and since it is computationally inefficient to compute the coefficients (without using the Leibnitz formulas in appendix E), it is best to use iterative methods to find eigenvalues associated directly with the matrix Eq.(2.131) instead of using Eq.(2.132) such as in the methods provided in MATLAB/LAPACK.

Chapter 3 A survey of noise estimation algorithms

3.1 Single channel noise estimation algorithms

This chapter presents the derivation of single-channel noise estimation algorithms. The algorithms rely on four observations that are presented in the excellent reference book by Loizou[8].

1. The power of the noisy signal decays often to the level of the noise in the subbands. We can then track the minimum of the noisy signal and make it correspond to an estimate of the noise PSD even during speech activity. Since the minimum is smaller than the average noise value, we will have to add a possibly adaptive bias factor to improve the estimate.
2. Silent segments in speech occur:
    - In the beginning and the end of a signal (voice) activity. This also occurs at the end of the plosive consonants.
    - During unvoiced fricatives at low frequencies generally below 2 kHz.
    - During voiced sounds above 4 kHz.

    Because of the different natures of the speech signal and the noise, the $iSNR[r,k]$ will be unevenly affected by the noise, in such a way that we can collect information about the noise power whenever the noised signal periodogram is not affected by speech. These observations lead to recursive-averaging noise estimation algorithms.
3. Similarly, by looking at the distribution of the power spectra along the frequency bands, we can notice that the most frequent value would be the noise level in most parts of the spectrum. This would mean that another way of estimating the minimum of the noise PSD is by looking for the first mode (most recurrent low power value) in each distribution for each frequency band. Typically for low frequencies, the distributions will have two modes one corresponding to the noise level and another to the speech level.
4. The low power levels occur much more frequently than the high energy levels that are caused mainly by the speech. For example in some frequency bands 80-90% of the power levels would be caused mainly by noise and low power components of the speech while the other 10-20% is caused by speech. The noise estimate is chosen as being the q^{th} quantile of the sorted power density values. For example, after sorting the power levels, we can choose the quantile number $q = 0.5$ which corresponds to the median.

These four observations led respectively to four algorithm classes:

1. Minimal-Tracking algorithms
2. Time-recursive averaging algorithms
3. Histogram-based algorithms
4. Quantile based algorithms

In [8] Loizou points out that usually the STFT spectra is used with 20-30 msec windows with 50% overlap between frames. Consecutive frames are then used to estimate the power spectrum described as an analysis segment. The time span of this segment can range from 400 msec to 1sec. The analysis segments need to be long enough to contain the low energy segments to have a good frequency resolution of the signal PSD but at the same time, it needs to be short enough(a good time-resolution) to track the fast changes of the noise PSD since it is non-stationary. This is the usual time-frequency resolution trade off choice.

For brevity, the chapter will omit quantile and histogram algorithms which rely on the last two of the four listed observations, and it will focus on the most popular approaches.

3.1.1 Minimal Tracking algorithms

Two algorithms of this type will be presented. The first algorithm is the Minimum statistics (MS) noise estimation algorithm developed by Martin [36, 37]. The second algorithm was proposed by Doblinger [38].

3.1.1.1 Minimum statistics (MS) noise estimation

Notation

The nomenclature for this algorithm is independent of the notation of the rest of the thesis.

$\sigma_y^2[r,k]$	Time-frequency dependent variance of noisy signal random variable
$f_{\lvert Y[r,k]\rvert^2}(t)$	Probability density function of spectrogram random variable
$u(t)$	Unit step function
$\sigma_s^2[r,k]$	Time-frequency dependent signal variance
$\sigma_\eta^2[r,k]$	Time-frequency dependent noise variance
$\mathcal{CN}(0,\sigma_y^2 I)$	Zero mean complex Gaussian function with variance σ_y^2
$\alpha[r,k]$	Smoothing factor
$P[r,k]$	Periodogram
$\alpha_{opt}[r,k]$	Optimal smoothing factor
$F_P(t)$	Cumulative distribution function of random variable P
P_{min}	Random variable representing the minimum sample of a set of periodograms
D	Number of samples in the set of periodograms
$\hat{P}$	Estimate of the periodogram
B_{min}	Bias factor
$Q_{eq}[r,k]$	Equivalent degrees of freedom
$H(D)$	Power term
$\Gamma(x)$	Gamma function (analytic continuation of $x-1!$ onto the real domain)
$M(D)$	Correction parameter
$var(P[r,k])$	Variance of $P[r,k]$
$\widehat{var}(P[r,k])$	Variance estimate of $P[r,k]$
$\bar{P}[r,k]$	Mean of $P[r,k]$
$\overline{P^2}[r,k]$	Second moment of $P[r,k]$
$\beta[r,k]$	Smoothing factor
U	Number of subwindows
V	Number of samples in each subwindow

Table 2: List of symbols for the MS algorithm

Principles

Let $y[n] = x[n] + \eta[n]$ be our sampled noisy speech signal which is a sum of a clean signal $x[n]$ and a noise signal $\eta[n]$. It is assumed $\eta[n]$ and $x[n]$ are statistically independent. The noisy speech signal is transformed into the discrete STFT domain by multiplying a frame of L_w consecutive samples of $y[n]$ with a window $w[n]$ and computing the FFT of length N_{fft}. Choosing $N_{fft} = L_w$ the discrete STFT is written as:

$$Y[r,k] = \sum_{m=0}^{N_{fft}-1} y[m+rR]w[m]e^{-\frac{j2\pi km}{N_{fft}}}. \quad \textbf{3.1}$$

From appendix A the instantaneous periodogram estimate probability density function has been derived under the condition that $Y[r,k]$ is a zero-mean complex gaussian distribution with radial symmetry, i.e. $Y[r,k] \sim \mathcal{CN}\left(0, \sigma_y^2 I\right)$ [39].

$$f_{|Y[r,k]|^2}(t) = \frac{1}{\sigma_y^2[r,k]} e^{-\frac{t}{\sigma_y^2[r,k]}} u(t), \quad \textbf{3.2}$$
$$u(t) = \begin{cases} 0; t \leq 0 \\ 1; t > 0' \end{cases}$$

where the statistical independence of the signal and noise imply that $\sigma_y^2[r,k] = \sigma_s^2[r,k] + \sigma_\eta^2[r,k]$,with $\mathcal{E}(|S[r,k]|^2) = \sigma_s^2[r,k]$, and $\mathcal{E}(|N[r,k]|^2) = \sigma_\eta^2[r,k]$. Because of the additivity of the variances, we can estimate $\sigma_\eta^2[r,k]$ by tracking the minimum of the periodogram of y, since $\sigma_y^2[r,k]$ often decays to $\sigma_\eta^2[r,k]$ in the subbands even during speech activity. The minimum of the noise periodogram tends to be lower than the true noise periodogram. As for the periodogram estimate, the following recursion is used:

$$P[r,k] = \alpha[r,k]P[r-1,k] + (1-\alpha[r,k])|Y[r,k]|^2. \quad \textbf{3.3}$$

$\alpha[r,k]$ is between 0 and 1. If $\alpha[r,k]$ is 1 then $P[r,k] = P[r-1,k]$ else if $\alpha[r,k] = 0$ then the periodogram is $P[r,k] = |Y[r,k]|^2$.

Derivation of the suboptimal smoothing factor

To derive the optimal $\alpha[r,k]$, speech pause is assumed and $P[r,k]$ is made as close as possible to $\sigma_\eta^2[r,k]$ by minimizing the conditional mean square error

$$\mathcal{J} = \mathcal{E}\left(\left(P[r,k] - \sigma_\eta^2[r,k]\right)^2 | P[r-1,k]\right), \quad \textbf{3.4}$$
$$\rightarrow \frac{\partial \mathcal{J}}{\partial \alpha[r,k]} = 0.$$

Solving for $\alpha[r,k]$, we obtain the optimal smoothing factor:

$$\alpha_{opt}[r,k] = \frac{1}{1 + \left(\frac{P[r-1,k]}{\sigma_\eta^2[r,k]} - 1\right)^2}. \quad \textbf{3.5}$$

For tracking error reasons [8, 37] $\alpha_{opt}[r,k]$ is modified to

$$\alpha[r,k] = \alpha_{max}\alpha_{opt}[r,k]\alpha_c[r,k] = \frac{\alpha_{max}\alpha_c[r,k]}{1+\left(\frac{P[r-1,k]}{\sigma_\eta^2[r,k]}-1\right)^2}, \qquad \alpha_{max} = 0.96, \tag{3.6}$$

with

$$\alpha_c[r,k] = 0.7\alpha_c[r-1,k] + 0.3\max(\tilde{\alpha}_c[r,k], 0.7),$$
$$\alpha_c[r,k] = \frac{1}{1+\left(\frac{\sum_{k=0}^{M-1}P[r-1,k]}{\sum_{k=0}^{M-1}|Y[r,k]|^2}-1\right)^2}. \tag{3.7}$$

For highly nonstationary noise, a SNR dependent variable lower limit α_{min} to $\alpha[r,k]$ is defined [37].

Derivation of bias factor

Because for nontrivial densities the minimum value of a set of random variables is smaller than their mean, we need to adjust the estimate with a bias factor B_{min}. Since periodograms are strictly positive, the average of the smallest value taken from D sample periodograms will be expressed by:

$$\mathcal{E}(P_{min}) = \int_0^\infty \left(1-F_P(t)\right)^D dt = \sigma_y^2[r,k]\int_0^\infty \left(1-F_{P|\sigma_y^2[r,k]=1}(t)\right)^D dt$$
$$= \sigma_y^2[r,k]\mathcal{E}\left(P_{min}|_{\sigma_y^2[r,k]=1}\right), \tag{3.8}$$

where $F_P(t)$ is the cumulative distribution of a sample periodogram P. The average $\mathcal{E}(P_{min})$ is proportional to $\sigma_y^2[r,k]$, and the average P would yield $\sigma_y^2[r,k]$ or,

$$\mathcal{E}(P) = \sigma_y^2[r,k]. \tag{3.9}$$

Multiplying the minimum with the bias factor and then taking the expected value yields,

$$\hat{P} = B_{min}P_{min},$$
$$\mathcal{E}(\hat{P}) = \mathcal{E}(P) = B_{min}\mathcal{E}(P_{min}),$$
$$B_{min}[r,k] = \frac{\mathcal{E}(P[r,k])}{\mathcal{E}(P_{min}[r,k])} = \frac{\sigma_y^2[r,k]}{\sigma_y^2[r,k]\mathcal{E}\left(P_{min}|_{\sigma_y^2[r,k]=1}\right)}$$
$$= \frac{1}{\mathcal{E}\left(P_{min}|_{\sigma_y^2[r,k]=1}\right)}. \tag{3.10}$$

In appendix A, the derivation of the pdf of a periodogram estimate is done under the conditions that the successive periodograms estimates are independent. In appendix C the characteristic function of the periodogram estimate is derived under the assumption that the periodogram bins are correlated. They are correlated in our case since there is a 50% overlap between the frames. It

is not trivial to derive a closed form expression for B_{min} even for uncorrelated frames. For this reason B_{min} was approximated using asymptotic results [37] by

$$B_{min}[r,k] \approx 1 + 2\frac{(D-1)}{\tilde{Q}_{eq}[r,k]}\Gamma\left(1 + \frac{2}{Q_{eq}[r,k]}\right)^{H(D)}, \quad \textbf{3.11}$$

where

$$\tilde{Q}_{eq}[r,k] = \frac{Q_{eq}[r,k] - 2M(D)}{1 - M(D)}. \quad \textbf{3.12}$$

Also named equivalent degrees of freedom, Q_{eq} is determined by

$$Q_{eq}[r,k] = \frac{2\sigma_\eta^4[r,k]}{var(P[r,k])}. \quad \textbf{3.13}$$

The approximate equivalent degrees of freedom for recursive and non-recursive periodogram distribution estimates are derived by Welch [40] for different types of windows. In [36], it is mentioned that for whatever window used, the distributions can be approximated with a chi-square distribution with an equivalent degrees of freedom parameter. It is mentioned in [37] that small values of Q_{eq} occur whenever a significant amout of speech power is present, and that it is unlikely that $P[r,k]$ attains a minimum in this case. Hence,

$$B_{min}\big(D, Q_{eq}[r,k]\big) \approx 1 + 2\frac{(D-1)}{\tilde{Q}_{eq}[r,k]} \quad \textbf{3.14}$$

can also be used. The estimation of the unbiased noise power based on minimum statistics is thus

$$\hat{\sigma}_\eta^2[r,k] \approx B_{min}\big(D, \hat{Q}_{eq}[r,k]\big)P_{min}[r,k]. \quad \textbf{3.15}$$

The equivalent degrees of freedom is estimated with

$$\hat{Q}_{eq}^{-1}[r,k] \approx \min\left(\frac{\widehat{var}(P[r,k])}{2\hat{\sigma}_\eta^4[r,k]}, 0.5\right). \quad \textbf{3.16}$$

The approximation of the first moment, second moments and the variance of $P[r,k]$ are done using

$$\bar{P}[r,k] = \beta[r,k]\bar{P}[r-1,k] + (1-\beta[r,k])P[r,k], \quad \textbf{3.17}$$

$$\overline{P^2}[r,k] = \beta[r,k]\overline{P^2}[r-1,k] + (1-\beta[r,k])P^2[r,k], \quad \textbf{3.18}$$

$$\widehat{var}(P[r,k]) = \overline{P^2}[r,k] - \bar{P}^2[r,k] \quad \textbf{3.19}$$

respectively.

The term $\beta[r,k]$ is computed with

$$\beta[r,k] = \alpha^2[r,k]. \quad \textbf{3.20}$$

A correction factor $\beta_c[r,k]$ can be multiplied to $\beta[r,k]$ to prevent $B_{min}\left(D,\hat{Q}_{eq}[r,k]\right)$ from pushing the minimum to values that are too small whenever the estimate of $P[r,k]$ has a large variance with

$$\beta_c[r,k] = 1 + 2.12\sqrt{\frac{1}{M}\sum_{k=0}^{M-1}\hat{Q}_{eq}^{-1}[r,k]}. \quad \textbf{3.21}$$

Searching for the minimum

On each frame of D consecutive periodograms we do $D-1$ comparisons to find the minimum. Because this implies $D-1$ operations per time frame per frequency bin, another method requiring fewer operations has been developed in [37]. This second method is done with the search windows having no overlap so we have $D-1$ compare operations per search frame per frequency bin. The problem with this is that we can have a delay of $2D$ samples. Another algorithm that lies between the two cases of maximal overlap and no overlap is done when the search frame is divided into U subwindows of V samples. Every V samples the minimum is updated and stored. The stored U minimums are then compared together to find the overall minimum of the search frame. Subsequently, the minimum of the first frame is discarded from the set and another minimum is added to it when the search frame advances of V samples. In other words the overlap of the search window is adjusted to $D-V$ samples, but this is done efficiently [37].

3.1.1.2 Continuous spectral minimum tracking

Another noise power estimator was developed by Doblinger where the periodogram is updated continuously like before with the usual recursive equation:

$$\hat{\Gamma}_{\mathrm{y}}[r,k] = \alpha\hat{\Gamma}_{\mathrm{y}}[r-1,k] + (1-\alpha)|Y[r,k]|^2, \quad \textbf{3.22}$$

where $\alpha \in [0.7,0.9]$.

The rule for estimating the minimum is:

if $\Gamma_\eta[r-1,k] < \hat{\Gamma}_{\mathrm{y}}[r,k]$

$$\hat{\Gamma}_\eta[r,k] = \gamma\hat{\Gamma}_\eta[r-1,k] + \frac{1-\gamma}{1-\beta}\left(\hat{\Gamma}_{\mathrm{y}}[r,k] - \beta\hat{\Gamma}_{\mathrm{y}}[r-1,k]\right) \quad \textbf{3.23}$$

else

$$\hat{\Gamma}_\eta[r,k] = \hat{\Gamma}_{\mathrm{y}}[r,k].$$

Typical parameters are $\alpha = 0.7$, $\beta = 0.96$, $\gamma = 0.998$. The parameters yield a noise adaptation of 0.2 to 0.4 seconds [38], so the estimator works well for an instantaneous noise power density estimator and the parameters can be modified to change the adaptation time. The term $\hat{\Gamma}_{\mathrm{y}}[r,k] - \beta\hat{\Gamma}_{\mathrm{y}}[r-1,k]$ in the estimator equation is approximately the discrete derivative of $\hat{\Gamma}_{\mathrm{y}}[r,k]$ since β is close to 1. Whenever there is a narrow and large sudden increase in the speech power spectrum the derivative term will be large. This induces an overestimation error in the noise periodogram estimate.

3.1.2 Time-recursive averaging algorithms

There are many algorithms of this type presented in [8], but the only one that will be presented is the one developed by Hendriks et al.[21] because of its simple solution.

3.1.2.1 MMSE based noise PSD tracking with low complexity

The new notation of the MMSE algorithm holds only for this section and not for the rest of the thesis.

R	Random variable representing the magnitude of noisy signal
N	Random variable modelling the additive noise
Θ	The phase random variable model of the noisy signal
W	The magnitude random variable model of the additive noise
σ	Phase random variable model of the noise
I_0	Modified Bessel function of the first kind of order 0
$\mathcal{M}("")$	Confluent hypergeometric function
γ	*A posteriori* SNR
ξ	*A priori* SNR
$B^{-1}(\xi)$	Inverse bias factor
$\widetilde{\sigma_W^2}$	Noise variance estimator
$\hat{\xi}_{ML}[r,k]$,	Maximum likelihood *a priori* SNR estimator
$\hat{\xi}_{DD}[r,k]$	Decision directed *a priori* SNR estimator

Table 3: List of symbols for the MMSE algorithm

Consider the usual problem $Y[m,k] = X[m,k] + N[m,k]$, where $X[m,k]$ is statistically independent of $N[m,k]$. Since all expressions are per time frame[8] m and frequency bin k, their dependence on them is omitted for clarity. If we express the noise and noisy speech discrete STFT in polar notation we get $Y = Re^{j\Theta}, N = We^{j\sigma}$. To estimate the noise PSD, an MMSE estimator of the noise magnitude squared DFT coefficients (W^2) is exploited. The MMSE estimator of W^2 is defined by the conditional expectation $\mathcal{E}(W^2|Y)$. In this section, uppercase letters denote random variables and lowercase ones denote their realizations. Using Bayes' rule:

$$\mathcal{E}(W^2|Y) = \frac{\int_0^\infty \int_0^{2\pi} w^2 f_{Y|W,\sigma}(y|w,\sigma) f_{W,\Psi}(n,\sigma) d\sigma dw}{\int_0^\infty \int_0^{2\pi} f_{Y|W,\sigma}(y|w,\sigma) f_{W,\Psi}(n,\sigma) d\sigma dw}. \quad \textbf{3.24}$$

Assuming both the speech and noise DFT coefficients to have a complex-Gaussian distribution, we get

$$f_{Y|W,\sigma}(y|w,\sigma) = \frac{1}{\pi\sigma_X^2} \exp\left(\frac{2wr\cos(\sigma-\theta) - r^2 - w^2}{\sigma_X^2}\right), \quad \textbf{3.25}$$

and

[8] Here the time frame is labelled with an *m* to avoid confusion with a realisation of the random variable *R* i.e. *r*.

$$f_{W,\Theta}(n,\sigma) = \frac{w}{\pi\sigma_W^2}\exp\left(-\frac{w^2}{\sigma_W^2}\right). \quad \textbf{3.26}$$

It follows that

$$\int_0^{2\pi} f_{Y|W,\Theta}(y|w,\sigma) f_{W,\Theta}(n,\sigma)d\sigma = \frac{w}{\pi^2\sigma_W^2\sigma_X^2} e^{-\frac{w^2}{\sigma_W^2}} e^{-\frac{r^2+w^2}{\sigma_X^2}} I_0\left(\frac{2rw}{\sigma_X^2}\right), \quad \textbf{3.27}$$

where I_0 is the modified Bessel function of order 0. Note that the above result does not depend on θ. Inserting the result of Eq.(3.27) into Eq.(3.24), $\mathcal{E}(W^2|Y)$ becomes

$$\mathcal{E}(W^2|Y) = \frac{\int_0^\infty w^3 e^{-\frac{w^2}{\sigma_W^2}} e^{-\frac{r^2+w^2}{\sigma_X^2}} I_0\left(\frac{2rw}{\sigma_X^2}\right)dw}{\int_0^\infty w e^{-\frac{w^2}{\sigma_W^2}} e^{-\frac{r^2+w^2}{\sigma_X^2}} I_0\left(\frac{2rw}{\sigma_X^2}\right)dw}. \quad \textbf{3.28}$$

Using the following formula

$$\int_0^\infty x^\nu e^{-ax^2} I_0(bx)dx = \frac{1}{2\sqrt{a^{\nu+1}}}\Gamma\left(\frac{\nu+1}{2}\right)\mathcal{M}\left(\frac{\nu+1}{2},1,\frac{b^2}{4a}\right), \quad \textbf{3.29}$$

where $\mathcal{M}(\text{""})$ is the confluent hypergeometric function[8, 16] we have

$$\mathcal{E}(W^2|Y) = \frac{\mathcal{M}\left(2,1,\frac{r^2}{\sigma_X^4}\left(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_W^2}\right)^{-1}\right)}{\left(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_W^2}\right)\mathcal{M}\left(1,1,\frac{r^2}{\sigma_X^4}\left(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_W^2}\right)^{-1}\right)}. \quad \textbf{3.30}$$

By expanding the function $\mathcal{M}(\text{""})$ the above expression can be simplified to

$$\mathcal{E}(W^2|Y) = \frac{\sigma_X^2\sigma_W^2}{\sigma_W^2+\sigma_X^2} + \frac{r^2\sigma_W^4}{(\sigma_W^2+\sigma_X^2)^2} = \left(\frac{\xi}{\gamma(1+\xi)} + \frac{1}{(1+\xi)^2}\right)r^2, \quad \textbf{3.31}$$

with the *a posteriori* SNR defined as

$$\gamma = \frac{r^2}{\sigma_W^2}, \quad \textbf{3.32}$$

and the *a priori* SNR

$$\xi = \frac{\sigma_X^2}{\sigma_W^2}. \quad \textbf{3.33}$$

Since the random variables X and N are assumed to be gaussian, Y has a complex Gaussian distribution of variance $\sigma_Y^2 = \sigma_X^2 + \sigma_W^2$. The expression $\mathcal{E}(W^2|Y)$ is independent of Θ and depends only on R which is Rayleigh distributed. We have $\mathcal{E}(W^2|Y) = \mathcal{E}(W^2|R)$ and

$$f_R(r) = 2\frac{r}{\sigma_Y^2} e^{-\frac{r^2}{\sigma_Y^2}}. \tag{3.34}$$

To check if $\mathcal{E}(W^2|R)$ is biased we take its expectation in terms of R.

$$\begin{aligned} \mathcal{E}_R\big(\mathcal{E}(W^2|R)\big) &= \int_0^\infty \mathcal{E}(W^2|R) f_R(r)\, dr \\ &= \int_0^\infty \left[\frac{\sigma_X^2\sigma_W^2}{\sigma_W^2+\sigma_X^2} + \frac{r^2\sigma_W^4}{(\sigma_W^2+\sigma_X^2)^2}\right] 2\frac{r}{\sigma_Y^2} e^{-\frac{r^2}{\sigma_Y^2}} dr = \sigma_W^2. \end{aligned} \tag{3.35}$$

The estimator $\mathcal{E}(W^2|R)$ is thus unbiaised. Since the *a priori* SNR is not known, the estimator used for it will introduce a bias to the noise power estimate. The maximum likelihood estimator for the *a priori* SNR developed by Ephraim and Malah [41] is

$$\hat{\xi}_{ML}[r,k] = \max(\gamma_{ML} - 1, 0) = \max\left(\frac{|Y[r,k]|^2}{\widehat{\sigma_W^2}[r-1,k]} - 1, 0\right). \tag{3.36}$$

The bias is defined as the ratio between the expected value of the theoretical estimator and the practical estimator:

$$B = \frac{\int_0^\infty \mathcal{E}(W^2|R) f_R(r) dr}{\int_0^\infty \mathcal{E}\big(W^2|R,\hat{\xi}_{ML}\big) f_R(r) dr} = \frac{\sigma_W^2}{\int_0^\infty \mathcal{E}\big(W^2|R,\hat{\xi}_{ML}\big) f_R(r) dr}. \tag{3.37}$$

To evaluate the denominator integral, we use $\hat{\xi} = (\gamma - 1)u(r - \sigma_W)$, where "$u$" is the unit step function:

$$\begin{aligned} \int_0^\infty \mathcal{E}\big(W^2|R,\hat{\xi}\big) f_R(r) dr &= \int_0^\infty \left(\frac{\hat{\xi}}{\gamma\big(1+\hat{\xi}\big)} + \frac{1}{\big(1+\hat{\xi}\big)^2}\right) r^2 f_R(r) dr \\ &= \int_0^{\sigma_W} r^2 f_R(r) dr + \sigma_W^2 \int_{\sigma_W}^\infty f_R(r) dr \\ &= \frac{2}{\sigma_Y^2}\int_0^{\sigma_W} r^3 e^{-\frac{r^2}{\sigma_Y^2}} dr + 2\frac{\sigma_W^2}{\sigma_Y^2}\int_{\sigma_W}^\infty \frac{r}{\sigma_Y^2} e^{-\frac{r^2}{\sigma_Y^2}} dr \\ = \sigma_Y^2 - e^{-\frac{\sigma_W^2}{\sigma_Y^2}}(\sigma_W^2 + \sigma_Y^2) + \sigma_W^2 e^{-\frac{\sigma_W^2}{\sigma_Y^2}} &= \sigma_W^2\frac{\sigma_Y^2}{\sigma_W^2}\left(1 - e^{-\frac{\sigma_W^2}{\sigma_Y^2}}\right) = \sigma_W^2(1+\xi)\left(1 - e^{-\frac{1}{1+\xi}}\right). \end{aligned} \tag{3.38}$$

The inverse bias is thus equal to

$$B^{-1}(\xi) = (1+\xi)\left(1 - e^{-\frac{1}{1+\xi}}\right). \quad \textbf{3.39}$$

Note that in [21], the above result is not in the same form, but it can be shown by expanding the incomplete gamma function found in the article that they are equivalent. Finally, the authors chose to use the ML estimator of ξ in Eq.(3.36) to compute $\mathcal{E}(W^2|R)$ and the decision directed approach SNR $\hat{\xi}_{DD}$[41] to compute the bias factor. Hence, the final estimator is thus

$$\widetilde{\sigma_W^2} = \mathcal{E}\left(W^2|R, \hat{\xi}_{ML}\right) B\left(\hat{\xi}_{DD}\right). \quad \textbf{3.40}$$

where,

$$\hat{\xi}_{DD}[r,k] = a\frac{\hat{X}^2[r-1,k]}{\widehat{\sigma_W^2}[r-1,k]} + (1-a)\max\left(\frac{|Y[r,k]|^2}{\widehat{\sigma_W^2}[r-1,k]} - 1{,}0\right), \quad \textbf{3.41}$$

$$a = 0.98.$$

In the above formula we need an estimate of the amplitude of X. The Wiener estimator of the signal amplitude is used to compute $\hat{X}$ with:

$$\hat{X}[r,k] = \frac{\hat{\xi}_{DD}[r,k]}{1+\hat{\xi}_{DD}[r,k]}|Y[r,k]|. \quad \textbf{3.42}$$

3.2 Two channel noise estimation algorithms

Three algorithms are presented here that discuss the topic from a dual channel perspective. All algorithms exploit a priori information about the diffuse noise coherence field. All algorithms assume that there is a single source, statistically independent from the noise. The first algorithm by Jeub et al.[1] adds the hypothesis that the transfer function for each channel is unity because of the assumption that the sources to microphones distances are smaller than the critical distance[9]. The algorithm by Kamkar-Parsi and Bouchard [3] makes the assumption that the noise coherence is known without knowing the vector transfer function. The last algorithm by Jeub et al.[2] exploits power level differences to distinguish if the signal in the second channel is affected by speech or not, supposing that the first channel has a direct transfer function for the source and that the source sound pressure is smaller in the second (noise reference) channel.

Since we have now two channels, it is natural to adopt the MIMO perspective notation. The time frame and frequency bin dependence is again dropped for brevity:

$$\mathbf{\Gamma}_y = \mathbf{\Gamma}_x + \mathbf{\Gamma}_\eta. \quad \textbf{3.43}$$

$\mathbf{\Gamma}_y, \mathbf{\Gamma}_x, \mathbf{\Gamma}_\eta$ are the power spectral density matrix of the noisy signal, the clean signal and the noise respectively. It is assumed for all 3 methods that the noise field is homogeneous and diffuse, so

[9] The critical distance is the distance in which the direct path sound energy equals the reverberant path sound energy.

consequently $\boldsymbol{\Gamma}_\eta$ is a product of the noise auto-PSD Γ_η and the coherence matrix $\boldsymbol{\Phi}$. The signals $x[n]$ are generally assumed to be from directional sources[10] consequently, $\boldsymbol{\Gamma}_x = \boldsymbol{H}_{M\times N}\boldsymbol{\Gamma}_s\boldsymbol{H}_{N\times M}^H = \boldsymbol{H}\boldsymbol{\Gamma}_s\boldsymbol{H}^H$. M is equal to 2(number of microphones), N equals the number of sources, $\boldsymbol{\Gamma}_s$ is the sources PSD matrix, $\boldsymbol{H}$ is the matrix of transfer functions and $\boldsymbol{A}^H$ denotes the conjugate transpose of a matrix $\boldsymbol{A}$. Since, for all cases a single source is assumed $\boldsymbol{H}$ is now a vector and $\boldsymbol{\Gamma}_x = \Gamma_s\boldsymbol{H}\boldsymbol{H}^H$.

3.2.1 Robust dual-channel noise PSD estimation [1]

In this method, the presence of one source is assumed and the source microphones distance is shorter than the critical distance. For practical reasons [1], $\boldsymbol{H} \approx \mathbf{1}_{2\times 1} = \mathbf{1}$. The sound measured at a microphone is a combination of reverberant and direct sound. The equation to solve is then

$$\boldsymbol{\Gamma}_y = \Gamma_s\mathbf{1}\mathbf{1}^T + \Gamma_\eta\boldsymbol{\Phi}. \tag{3.44}$$

The diagonal terms share the same algebraic equations:

$$\begin{aligned}\Gamma_{y_1} &= \Gamma_s + \Gamma_\eta,\\ \Gamma_{y_2} &= \Gamma_s + \Gamma_\eta.\end{aligned} \tag{3.45}$$

Then by taking the geometrical mean of the diagonal terms:

$$\sqrt{\Gamma_{y_1}\Gamma_{y_2}} = \Gamma_s + \Gamma_\eta. \tag{3.46}$$

The third and last equations come in pair of complex conjugates

$$\begin{aligned}\Gamma_{y_1y_2} &= \Gamma_s + \Gamma_\eta\Phi_{\eta_1\eta_2},\\ \Gamma_{y_1y_2}^* &= \Gamma_s + \Gamma_\eta\Phi_{\eta_1\eta_2}^*.\end{aligned} \tag{3.47}$$

Then combining the off diagonal term equations:

$$\begin{aligned}\left(\Gamma_{y_1y_2}^* + \Gamma_{y_1y_2}\right) &= 2\Gamma_s + \Gamma_\eta\left(\Phi_{\eta_1\eta_2} + \Phi_{\eta_1\eta_2}^*\right),\\ Re\left(\Gamma_{y_1y_2}\right) &= \Gamma_s + \Gamma_\eta Re\left(\Phi_{\eta_1\eta_2}\right).\end{aligned} \tag{3.48}$$

Combining the off-diagonal term equation and the diagonal term equation we get

$$Re\left(\Gamma_{y_1y_2}\right) - \Gamma_\eta Re\left(\Phi_{\eta_1\eta_2}\right) = \sqrt{\Gamma_{y_1}\Gamma_{y_2}} - \Gamma_\eta. \tag{3.49}$$

Solving for Γ_η gives

$$\frac{\sqrt{\Gamma_{y_1}\Gamma_{y_2}} - Re\left(\Gamma_{y_1y_2}\right)}{1 - Re\left(\Phi_{\eta_1\eta_2}\right)} = \Gamma_\eta. \tag{3.50}$$

[10] The sources here include both directional interferers and targets.

3.2.2 Binaural approach [3]

The algorithm will be briefly explained here. The main assumption in this technique is that there is only one source and that the coherence matrix of the noise field is known. As usual, the source and the homogeneous diffuse noise are independent. We thus have the following equation to solve

$$\boldsymbol{\Gamma}_y = \Gamma_s \boldsymbol{H}\boldsymbol{H}^H + \Gamma_\eta \boldsymbol{\Phi},$$

or

$$\begin{bmatrix} \Gamma_{y_1} & \Gamma_{y_1 y_2} \\ \Gamma^*_{y_1 y_2} & \Gamma_{y_2} \end{bmatrix} = \Gamma_s \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}^H + \Gamma_\eta \begin{bmatrix} 1 & \Phi_{\eta_1 \eta_2} \\ \Phi_{\eta_1 \eta_2} & 1 \end{bmatrix}, \quad \textbf{3.51}$$

where H_i is the transfer function from the source to the i[th] microphone. First the Wiener filter is defined as

$$H_W = \frac{\Gamma_{y_1 y_2}}{\Gamma_{y_2}}. \quad \textbf{3.52}$$

An extra step for computing the autocorrelation quantity γ_e is computed in the time domain

$$\gamma_e[n] = \mathcal{E}(e[n] e^*[n-i]),$$

with

$$e[n] = y_1[n] - y_2[n] * h_w[n], \quad \textbf{3.53}$$

where " $*$ " denotes the discrete time linear convolution. In the Fourier domain, the PSD of the residual error $e[n]$ is theoretically

$$\Gamma_e = \Gamma_{y_1} - \Gamma_{y_2} |H_W|^2. \quad \textbf{3.54}$$

It can be shown [3] using the diagonal terms of the matrix equation that the magnitude square of the Wiener filter in the frequency domain is,

$$|H_W|^2 = \frac{\left(\Gamma_{y_1} - \Gamma_\eta\right)\left(\Gamma_{y_2} - \Gamma_\eta\right) + \Phi^2 \Gamma_\eta^2 + \Gamma_{\mathrm{A}}}{\Gamma_{y_2}^2}, \quad \textbf{3.55}$$

where

$$\Gamma_{\mathrm{A}} = 2\Phi \Gamma_\eta \Gamma_s Re(H_1 H_2^*).$$

In [3] the solution of above quadratic equation for $\Gamma_\eta$with the negative root is chosen and the reason for why the negative root is the correct one is also explained. The solution is given by

$$\Gamma_\eta = \frac{\Gamma_{y_1} + \Gamma_{y_2} - 2\Phi \Gamma_{y_2} Re(H_W) \pm \Gamma_{root}}{2\left(1 - \Phi^2_{\eta_1 \eta_2}\right)},$$

$$\Gamma_{root} = \sqrt{2\Phi_{\eta_1 \eta_2} \Gamma_{y_2} Re(H_W) - 4\left(1 - \Phi^2_{\eta_1 \eta_2}\right)\Gamma_e \Gamma_{y_2} - \left(\Gamma_{y_1} + \Gamma_{y_2}\right)}. \quad \textbf{3.56}$$

The algorithm can be simplified by skipping the time domain processing and by noticing that the quadratic equation is the generalized characteristic polynomial of the following matrix binomial:

$$\mathbf{\Gamma}_y - \Gamma_\eta \mathbf{\Phi}. \tag{3.57}$$

Further details on the simplifications for this method will be discussed in Chapter 4.

3.2.3 Power Level Difference Noise Estimator (PLDNE) [2]

The paper in [2] explains how it is possible to get an estimator from the assumption that there is a sufficient attenuation of the desired speech signal between the 2 microphones, for example 10dB. In the first step, the normalized difference of the PSD $0 \leq \Delta\Gamma_{PLDNE}[r,k] \leq 1$ of the noisy input signal is computed for every time frame r and frequency bin k with,

$$\Delta\Gamma_{PLDNE}[r,k] = \left|\frac{\Gamma_{y_1}[r,k] - \Gamma_{y_2}[r,k]}{\Gamma_{y_1}[r,k] + \Gamma_{y_2}[r,k]}\right|, \tag{3.58}$$

and $\mathbf{\Gamma}_y$ is computed using Eq.(2.139) with $\alpha = \alpha_1$.

In case of background noise-only periods, $\Delta\Gamma_{PLDNE}$ will be close to zero if the input powers are equal. If $\Delta\Gamma_{PLDNE}$ is below a threshold ϕ_{min} the noise PSD estimate is determined directly from the signal $y_1[n]$ with

$$\hat{\Gamma}_\eta[r,k] = \alpha_2 \hat{\Gamma}_\eta[r-1,k] + (1-\alpha_2)|Y_1[r,k]|^2 ,$$

$$\text{if} \quad \Delta\Gamma_{PLDNE}[r,k] < \phi_{min}. \tag{3.59}$$

When there is no noise, the PSD inequality $\Gamma_{y1} > \Gamma_{y2}$ holds. Consequently, $\Delta\Gamma_{PLDNE}[r,k] \approx 1$. Thus α_2 will be set to 1, or equivalently

$$\hat{\Gamma}_\eta[r,k] = \hat{\Gamma}_\eta[r-1,k] ,$$

$$\text{if} \quad \Delta\Gamma_{PLDNE}[r,k] > \phi_{max}. \tag{3.60}$$

In between both cases, the noise PSD approximation is done by using the second input because the attenuated speech component in $y_2[n]$ can be neglected, i.e.

$$\hat{\Gamma}_\eta[r,k] = \alpha_3 \hat{\Gamma}_\eta[r-1,k] + (1-\alpha_3)|Y_2[r,k]|^2 ,$$

$$\text{if} \quad \phi_{min} < \Delta\Gamma_{PLDNE}[r,k] < \phi_{max}. \tag{3.61}$$

In case of babble noise presence, the PLDNE algorithm can be combined with other algorithms[2], for example[1, 21, 37].

If the noise is no longer homogeneous e.g., when an interfering talker is present, the power level difference is implemented with:

$$\Delta\Gamma_{PLDNE}[r,k] = \max\big(\Gamma_{y_1}[r,k] - \Gamma_{y_2}[r,k], 0\big). \tag{3.62}$$

To summarize the noise estimation based on $\Delta\Gamma_{PLDNE}$ we present the following table:

Presence		Order relation[11]		
Signal	**Noise**	**of Γ_{y_1} and Γ_{y_2}**	**Range of $\Delta\Gamma_{PLDNE}$**	$\hat{\Gamma}_{\eta}[r,k]$
No	Yes	$\Gamma_{y_1} \approx \Gamma_{y_2}$	$\Delta\Gamma_{PLDNE}[r,k] < \phi_{min}$	$\alpha_2\hat{\Gamma}_{\eta}[r-1,k] + (1-\alpha_2)\lvert Y_1[r,k]\rvert^2$
Yes	Yes	$\Gamma_{y_1} > \Gamma_{y_2}$	$\phi_{min} < \Delta\Gamma_{PLDNE}[r,k] < \phi_{max}$	$\alpha_3\hat{\Gamma}_{\eta}[r-1,k] + (1-\alpha_3)\lvert Y_2[r,k]\rvert^2$
Yes	No	$\Gamma_{y_1} > \Gamma_{y_2}$	$\phi_{max} < \Delta\Gamma_{PLDNE}[r,k]$	$\hat{\Gamma}_{\eta}[r-1,k]$

Table 4: PLDNE update equations summary

[11] In the noise-only case, Γ_{y_1} and Γ_{y_2} may not be of the same order of magnitude if the noise field is inhomogeneous.

Chapter 4 The proposed algorithms (★)

It is often important to know the noise PSD $\mathbf{\Gamma}_\eta$ because it is required for so many algorithms, e.g., if we first start with the objective of finding the MMSE filter that best estimates $\boldsymbol{X}$ given that we know $\boldsymbol{Y}$ and $\mathbf{\Phi}$, we will notice that we need the noise PSD. Plus, definitions of the SNR and other performance measures necessarily depend on $\mathbf{\Gamma}_\eta$. The problem is that we typically don't know the noise PSD matrix $\mathbf{\Gamma}_\eta$. This chapter will introduce new coherence based algorithms to estimate the noise PSD matrix from the MIMO perspective. For multichannel noise PSD matrix estimation, the first step consists of computing the auto and cross periodograms by using Eqs.(2.139, 2.140). It is important to use enough samples to ensure that $\mathbf{\Gamma}_y[r,k]$ has a full rank or $L > M$. But at the same time, we need to use L or α as small as possible to have a good time resolution and adapt quickly to non-stationary environments. In the algorithms, there is a need in computing generalized eigenpairs and the signal dimensionality. The signal dimensionality can be estimated as being fixed (e.g. $\widehat{N} = 1$), or it can be estimated using the AIC criterion discussed in section 2.8.

4.1 Known coherence matrix $\mathbf{\Phi}$

Eq.(2.112) can be rewritten as

$$\mathbf{\Phi}^{-1/2}\hat{\mathbf{\Gamma}}_y\mathbf{\Phi}^{-1/2} - \Gamma_\eta \boldsymbol{I} = \mathbf{\Phi}^{-\frac{1}{2}}\mathbf{\Gamma}_x\mathbf{\Phi}^{-\frac{1}{2}}. \tag{4.1}$$

The above can be seen as a noise whitening process.

Suppose that $\mathbf{\Phi}^{-1/2}\hat{\mathbf{\Gamma}}_y\mathbf{\Phi}^{-1/2}$ has the following eigendecomposition

$$\mathbf{\Phi}^{-1/2}\hat{\mathbf{\Gamma}}_y\mathbf{\Phi}^{-1/2} = \boldsymbol{V}\mathbf{\Lambda}\boldsymbol{V}^H,$$

with

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_\mathrm{M}), \tag{4.2}$$

and

$$\lambda_1 \geq \cdots \geq \lambda_\mathrm{M},$$

where $\boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{I}$ and $\boldsymbol{V}$ contains the column eigenvectors i.e., $\boldsymbol{V} = [\boldsymbol{v}_1|\cdots|\boldsymbol{v}_M]$. If $rank(\boldsymbol{\Gamma}_x) = N$, then $\mathbf{\Phi}^{-1/2}\hat{\mathbf{\Gamma}}_y\mathbf{\Phi}^{-1/2}$ has the following decomposition:

$$\mathbf{\Phi}^{-1/2}\hat{\mathbf{\Gamma}}_y\mathbf{\Phi}^{-1/2} = \begin{bmatrix}\boldsymbol{V}_x & \boldsymbol{V}_\eta\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_\mathrm{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_\eta\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H \\ \boldsymbol{V}_\eta^H\end{bmatrix},$$

with

$$\mathbf{\Lambda}_\mathrm{x} = \mathrm{diag}(\lambda_1, \dots, \lambda_\mathrm{N}), \qquad \mathbf{\Lambda}_\eta = \mathrm{diag}(\lambda_{\mathrm{N}+1}, \dots, \lambda_\mathrm{M}), \tag{4.3}$$

and

$$\boldsymbol{V}_x = [\boldsymbol{v}_1|\cdots|\boldsymbol{v}_N], \qquad \boldsymbol{V}_\eta = [\boldsymbol{v}_{N+1}|\cdots|\boldsymbol{v}_M].$$

The matrices $\boldsymbol{V}_x$ and $\boldsymbol{V}_\eta$ are the signal and noise subspace respectively. The eigenvectors have the following relations:

$$\boldsymbol{V}^H\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_x^H\boldsymbol{V}_x & \boldsymbol{V}_x^H\boldsymbol{V}_\eta \\ \boldsymbol{V}_\eta^H\boldsymbol{V}_x & \boldsymbol{V}_\eta^H\boldsymbol{V}_\eta \end{bmatrix} = \boldsymbol{I}_{M\times M} = \begin{bmatrix} \boldsymbol{I}_{N\times N} & \boldsymbol{0}_{N\times M-N} \\ \boldsymbol{0}_{M-N\times N} & \boldsymbol{I}_{M-N\times M-N} \end{bmatrix},$$
$$\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{V}_x\boldsymbol{V}_x^H + \boldsymbol{V}_\eta\boldsymbol{V}_\eta^H = \boldsymbol{I}_{M\times M}. \quad \textbf{4.4}$$

Using the maximum likelihood estimate of the noise auto-PSD estimate of Γ_η in Eq.(2.143)

$$\boldsymbol{\Phi}^{-1/2}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Phi}^{-1/2} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I} = \boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-\frac{1}{2}}. \quad \textbf{4.5}$$

Even if we know the true $\boldsymbol{\Phi}$ related to the environment, the instantaneous coherence matrix $\boldsymbol{\Phi}'$ will not be constant over time. Hence, Eq.(4.1) becomes

$$\boldsymbol{\Phi}^{-1/2}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Phi}^{-1/2} - \Gamma_\eta\left(\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{\Phi}'\boldsymbol{\Phi}^{-\frac{1}{2}}\right) = \boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{\Gamma}_x\boldsymbol{\Phi}^{-\frac{1}{2}}. \quad \textbf{4.6}$$

The eigenvalues are reliable if $\left(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Phi}'\boldsymbol{\Phi}^{-1/2}\right)$ is close to $\boldsymbol{I}$. The sensitivity issues related to the eigenvalue perturbation has been discussed in section 2.6. The one above relates specifically to the perturbation of $\boldsymbol{\Phi}'$. If the coherence matrix $\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Phi}'\boldsymbol{\Phi}^{-1/2}$ isn't close to $\boldsymbol{I}$ we get wrong estimates of Γ_η and consequently the rank estimate of our signal subspace will be erroneous. The risk that $\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Phi}'\boldsymbol{\Phi}^{-1/2}$ differs from $\boldsymbol{I}$ increases if $\boldsymbol{\Phi}$ is ill-conditioned (this can happen if at least two sensors are closely spaced). An ill-conditioned matrix has the characteristic that a small perturbation of the matrix will lead to large errors in the matrix function of the inversion type. We can thus expect the ML estimator of *N* to fail in the sense that the signal subspace dimensionality will be estimated larger and thus lead to underestimated values of $\hat{\Gamma}_\eta$, not to mention that the estimated noise subspace eigenvalues might be subject to errors too.

4.1.1 First estimate: Approximating the noise PSD matrix

One first estimate of the matrix $\boldsymbol{\Gamma}_\eta$ would be using a modified version of Eq.(2.134) where the solution is the average of the minimal eigenvalues (the average of the eigenvalues associated with the noise subspace):

$$\hat{\boldsymbol{\Gamma}}_\eta^{(1)} = \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Phi}. \quad \textbf{4.7}$$

It can be shown that the method presented in [3] is a special case of the above estimate where $M = 2$ and $\hat{N}$ is fixed at 1. One advantage[12] of this method is that there is no need in computing eigenvalues iteratively (for M smaller than 5) since we have developed closed form solutions for theses in section 0. Because of the good results in [3] we deduce that a fixed value of $\hat{N} = 1$ is a good estimate for M=2. However, when M increases it will become less likely that $\hat{N} = 1$ is the correct signal dimension if the number of speakers is of the order of M. This might lead to an increased error probability.

[12] As discussed earlier, it depends on how the roots are implemented.

The above estimate will always give noise PSDs results as if they came from a homogeneous noise field, i.e. the auto-PSDs are identical. That is not always true and there is a need in estimating inhomogeneous noise field noise PSDs even if we are given a noise coherence matrix. A binaural setting will almost certainly give different noise PSDs at each channel. Secondly we know that if there is no sources then $\hat{\mathbf{\Gamma}}_\eta = \hat{\mathbf{\Gamma}}_y$, but the above estimate doesn't take this into account.

4.1.2 Second estimate: Noise subspace update

Another way to solve this estimation issue is using projection matrices defined as

$$\begin{aligned} \boldsymbol{P}_\eta &= \boldsymbol{V}_\eta\left(\boldsymbol{V}_\eta^H\boldsymbol{V}_\eta\right)^{-1}\boldsymbol{V}_\eta^H = \boldsymbol{P}_{x^\perp} = \boldsymbol{I} - \boldsymbol{P}_x, \\ \boldsymbol{P}_x &= \boldsymbol{V}_x(\boldsymbol{V}_x^H\boldsymbol{V}_x)^{-1}\boldsymbol{V}_x^H = \boldsymbol{P}_{\eta^\perp} = \boldsymbol{I} - \boldsymbol{P}_\eta. \end{aligned} \quad \textbf{4.8}$$

The orthonormality of the eigenvectors shown in Eq.(4.4) allows simplification of the projection matrices:

$$\begin{aligned} \boldsymbol{P}_\eta &= \boldsymbol{V}_\eta\boldsymbol{V}_\eta^H = \boldsymbol{P}_{x^\perp} = \boldsymbol{I} - \boldsymbol{P}_x, \\ \boldsymbol{P}_x &= \boldsymbol{V}_x\boldsymbol{V}_x^H = \boldsymbol{P}_{\eta^\perp} = \boldsymbol{I} - \boldsymbol{P}_\eta. \end{aligned} \quad \textbf{4.9}$$

The estimate of $\boldsymbol{\Gamma}_\eta$ can be further improved by updating the instantaneous noise subspace:

$$\hat{\boldsymbol{\Gamma}}_\eta^{(2)} = \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{P}_\eta\boldsymbol{\Phi}^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{P}_\eta + \boldsymbol{P}_x\boldsymbol{\Phi}^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_\eta^{(1)}\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{P}_x\right)\boldsymbol{\Phi}^{\frac{1}{2}}$$

$$\hat{\boldsymbol{\Gamma}}_\eta^{(2)} = \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{P}_\eta\boldsymbol{\Phi}^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{P}_\eta + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{P}_x\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{\Phi}\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{P}_x\right)\boldsymbol{\Phi}^{\frac{1}{2}}$$

which simplified into

$$\begin{aligned} \hat{\boldsymbol{\Gamma}}_\eta^{(2)} &= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{P}_\eta\boldsymbol{\Phi}^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{P}_\eta + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{P}_x\right)\boldsymbol{\Phi}^{\frac{1}{2}} \\ &= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{V}_\eta\boldsymbol{\Lambda}_\eta\boldsymbol{V}_\eta^H + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{P}_x\right)\boldsymbol{\Phi}^{\frac{1}{2}} \\ &= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{V}_\eta\boldsymbol{\Lambda}_\eta\boldsymbol{V}_\eta^H + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}(\boldsymbol{I}-\boldsymbol{P}_\eta)\right)\boldsymbol{\Phi}^{\frac{1}{2}} \end{aligned} \quad \textbf{4.10}$$

$$= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{V}_\eta\left(\boldsymbol{\Lambda}_\eta - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\left(\boldsymbol{V}_\eta^H\boldsymbol{V}_\eta\right)^{-1}\right)\boldsymbol{V}_\eta^H + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I}\right)\boldsymbol{\Phi}^{\frac{1}{2}}$$

$$= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\boldsymbol{V}_\eta\left(\boldsymbol{\Lambda}_\eta - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I}_{M-\hat{N}\times M-\hat{N}}\right)\boldsymbol{V}_\eta^H + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I}\right)\boldsymbol{\Phi}^{\frac{1}{2}}$$

or

$$\hat{\boldsymbol{\Gamma}}_\eta^{(2)} = \boldsymbol{\Phi}^{\frac{1}{2}}\begin{bmatrix}\boldsymbol{V}_x & \boldsymbol{V}_\eta\end{bmatrix}\begin{bmatrix}\frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\eta\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H \\ \boldsymbol{V}_\eta^H\end{bmatrix}\boldsymbol{\Phi}^{\frac{1}{2}}$$

$$= \boldsymbol{\Phi}^{\frac{1}{2}}\left(\frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\widehat{N}}\boldsymbol{P}_x + \boldsymbol{V}_\eta\boldsymbol{\Lambda}_\eta\boldsymbol{V}_\eta^H\right)\boldsymbol{\Phi}^{\frac{1}{2}}$$

Eq.(4.10) are the steps used to compute $\hat{\boldsymbol{\Gamma}}_n$ by keeping the contribution of $\frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\widehat{N}}\boldsymbol{\Phi}$ in the signal subspace and updating the noise subspace with $\boldsymbol{V}_\eta\boldsymbol{\Lambda}_\eta\boldsymbol{V}_\eta^H$ from $\hat{\boldsymbol{\Gamma}}_y$. Note that the above formula has the following special cases:

$$\widehat{\boldsymbol{\Gamma}}_\eta^{(2)} = \hat{\boldsymbol{\Gamma}}_y; \widehat{N} = 0$$

and **4.11**

$$\widehat{\boldsymbol{\Gamma}}_\eta^{(2)} = \lambda_{min}\boldsymbol{\Phi} = \widehat{\boldsymbol{\Gamma}}_\eta^{(1)}; \widehat{N} = 1, M = 2.$$

We see above that when $\widehat{N} = 1, M = 2$ the second estimate of Eq.(4.19) equals the first estimate in Eq.(4.7). And as previously mentioned, the two-channel algorithm in section 3.2.2(Binaural approach [3]) is equivalent conceptually to the case when $\widehat{N} = 1, M = 2$. This has the consequence that when $\widehat{N} = 1, M = 2$, we can't find different noise power levels in the two channels, even though for a binaural setting $\widehat{\boldsymbol{\Gamma}}_\eta^{(2)} = \hat{\boldsymbol{\Gamma}}_y$ will have different noise PSDs for each channel. We can only figure them both when there is no active source, i.e., $\widehat{N} = 0$.

4.1.3 Third estimate: Coherence matrix update

From the noise subspace update in Eq.(4.19), we can estimate the instantaneous coherence with

$$\widehat{\boldsymbol{\Phi}}'[r,k] = \left(\hat{\boldsymbol{\Gamma}}_\eta[r,k] \circ \boldsymbol{I}\right)^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_\eta[r,k]\left(\hat{\boldsymbol{\Gamma}}_\eta[r,k] \circ \boldsymbol{I}\right)^{-\frac{1}{2}}. \quad \textbf{4.12}$$

This coherence sample permits an adaptive update of the coherence matrix itself, which is very useful since coherences are dependent on the environment. This is the way we can reduce the coherence matrix error with the following

$$\widehat{\boldsymbol{\Phi}}[r,k] = \mu\widehat{\boldsymbol{\Phi}}[r-1,k] + (1-\mu)\widehat{\boldsymbol{\Phi}}'[r,k],$$

with initial condition **4.13**

$$\widehat{\boldsymbol{\Phi}}[-1,k] = \boldsymbol{\Phi}.$$

The smoothing parameter μ needs to be close to one to have a good estimate of the coherence but at the same time if μ is too close to one, the adaptation time will be slow. Writing $\widehat{\boldsymbol{\Phi}}[r-1,k] = \widehat{\boldsymbol{\Phi}}_{r-1}$ for short, the final adaptive algorithm is:

$$\widehat{\boldsymbol{\Gamma}}_\eta^{(3)} = \widehat{\boldsymbol{\Phi}}_{r-1}^{1/2}\left(\boldsymbol{P}_\eta\widehat{\boldsymbol{\Phi}}_{r-1}^{-1/2}\hat{\boldsymbol{\Gamma}}_y[r,k]\widehat{\boldsymbol{\Phi}}_{r-1}^{-1/2}\boldsymbol{P}_\eta + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\widehat{N}}\boldsymbol{P}_x\right)\widehat{\boldsymbol{\Phi}}_{r-1}^{1/2}. \quad \textbf{4.14}$$

An advantage of this technique would be that we could use any reasonable initial condition ,e.g.$\widehat{\boldsymbol{\Phi}}[-1,k] = \boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{I}$. A disadvantage would be that if α is too small then estimating the signal subspace using the AIC becomes less reliable and so, there is a possibility that $\widehat{\boldsymbol{\Phi}}$ might actually diverge from the environments true coherence matrix. Underestimating or overestimating the dimensionality of the signal subspace will lead to overestimated or underestimated noise power levels respectively. This technique requires reliable estimates of $\widehat{N}$ or else the instantaneous coherence estimator might actually converge to the coherence of the signal subspace.

Unfortunately estimating the ML value $\widehat{N}$ relies on knowledge of $\mathbf{\Phi}$ itself. This would lead to completely unreliable estimates since to estimate $\widehat{\mathbf{\Phi}}$ we need a reliable $\widehat{N}$, but a reliable $\widehat{N}$ needs a good $\widehat{\mathbf{\Phi}}$ estimate. One possibility to remedy this situation would be to develop a new estimation scheme for $\widehat{N}$ that would be independent on $\mathbf{\Phi}$. In order to prove the validity of Eq.(4.14) numerically, we will compute a reliable $\widehat{N}$ that satisfies the following equation:

$$\widehat{N}[r,k] = \arg\min_{N} \sum_{i} \left| 10\log_{10}\left[\frac{\Gamma_{\eta_i}[r,k]}{\hat{\Gamma}_{\eta_i}[r,k,N]}\right]\right|; N \in [0, M-1],$$

$$\hat{\mathbf{\Gamma}}_{\eta}[r,k,N] = \mathbf{\Phi}^{\frac{1}{2}}\left(\boldsymbol{V}_{\eta}\left(\mathbf{\Lambda}_{\eta} - \frac{\mathrm{tr}(\mathbf{\Lambda}_{\eta})}{M-N}\boldsymbol{I}_{M-N\times M-N}\right)\boldsymbol{V}_{\eta}^{H} + \frac{\mathrm{tr}(\mathbf{\Lambda}_{\eta})}{M-N}\boldsymbol{I}\right)\mathbf{\Phi}^{\frac{1}{2}}. \quad \textbf{4.15}$$

In other words, we pick *N* such that it minimizes the log-error cost function. This quantity is not available in practice, however it is used nevertheless to show that if one has a good estimate of *N*, an adaptive estimate for $\mathbf{\Phi}$ becomes possible. Thus the algorithm to compute $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(3)}$ cannot yet be used in practice, since it uses a quantity not available.

Summarizing the 3 proposed algorithms of this section

Here is a table that summarizes the algorithms with $\alpha = 0.7$:

Step 1 Compute $\widehat{N}$ with the AIC in Eq.(2.146), or simply fix $\widehat{N}$ at 1.

Step 2 Compute $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)} = \frac{\mathrm{tr}(\mathbf{\Lambda}_{\eta})}{M-\widehat{N}}\mathbf{\Phi}$ with eigendecomposition described in Eq.(4.3).

Step 3 Update the noise subspace of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$with $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)} = \mathbf{\Phi}^{\frac{1}{2}}\left(\frac{\mathrm{tr}(\mathbf{\Lambda}_{\eta})}{M-\widehat{N}}\boldsymbol{P}_{x} + \boldsymbol{V}_{\eta}\mathbf{\Lambda}_{\eta}\boldsymbol{V}_{\eta}^{H}\right)\mathbf{\Phi}^{\frac{1}{2}}$.

Step 4 Update the coherence with Eq.(4.12) and Eq.(4.13).

Table 5: Noise PSD matrix estimation equations summary.

Steps 1 and 2 corresponds to the first algorithm. The second and third algorithm correspond to steps 3 and 4 respectively.

Noise reduction link

It is interesting to note that if we use Eq.(4.10) and the Wiener filter relation of Eq.(2.96) we get

$$\begin{aligned}
&\boldsymbol{\Gamma}_y \boldsymbol{W} = \boldsymbol{\Gamma}_x, \\
&\rightarrow \boldsymbol{\Gamma}_y \boldsymbol{W} = \boldsymbol{\Gamma}_y - \hat{\boldsymbol{\Gamma}}_\eta, \\
&\rightarrow \boldsymbol{\Phi}^{\frac{1}{2}}\begin{bmatrix}\boldsymbol{V}_x & \boldsymbol{V}_\eta\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{\mathrm{x}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\eta}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H \\ \boldsymbol{V}_\eta^H\end{bmatrix}\boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{W} = \boldsymbol{\Phi}^{\frac{1}{2}}\begin{bmatrix}\boldsymbol{V}_x & \boldsymbol{V}_\eta\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{\mathrm{x}} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H \\ \boldsymbol{V}_\eta^H\end{bmatrix}\boldsymbol{\Phi}^{\frac{1}{2}}, \\
&\rightarrow \boldsymbol{W} = \boldsymbol{\Phi}^{-\frac{1}{2}}\begin{bmatrix}\boldsymbol{V}_x & \boldsymbol{V}_\eta\end{bmatrix}\begin{bmatrix}\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H \\ \boldsymbol{V}_\eta^H\end{bmatrix}\boldsymbol{\Phi}^{\frac{1}{2}}, \\
&= \boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{V}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)\boldsymbol{V}_x^H\boldsymbol{\Phi}^{\frac{1}{2}}.
\end{aligned} \tag{4.16}$$

The covariance matrix of the Wiener solution would be then given by:

$$\begin{aligned}
&\mathcal{E}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^H\right) = \mathcal{E}(\boldsymbol{W}^H\boldsymbol{Y}\boldsymbol{Y}^H\boldsymbol{W}) \\
&= \boldsymbol{\Gamma}_x\boldsymbol{\Gamma}_y^{-1}\mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}^H)\boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_x = \boldsymbol{\Gamma}_x\boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_x.
\end{aligned} \tag{4.17}$$

We see that $\mathcal{E}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^H\right) \neq \boldsymbol{\Gamma}_x$.To obtain a covariance matrix of the estimated input that equals the covariance matrix $\boldsymbol{\Gamma}_x$ we can extend the spectral subtraction with the MIMO perspective by using the square root of the Wiener matrix. This can be found easily due to the similarity of $\boldsymbol{W}$ and $\boldsymbol{V}\begin{bmatrix}\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{V}^H$ with respect to the similarity transformation $\boldsymbol{\Phi}^{\frac{1}{2}}$:

$$\boldsymbol{W}^{\frac{1}{2}} = \boldsymbol{G} = \boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{V}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)^{1/2}\boldsymbol{V}_x^H\boldsymbol{\Phi}^{\frac{1}{2}}. \tag{4.18}$$

Note that the diagonal terms in the matrix $\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}$ are positive since eigenvalues are ordered in decreasing order.

The covariance matrix of this spectral subtraction solution would be then given by:

$$\begin{aligned}
&\mathcal{E}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^H\right) = \mathcal{E}\left(\boldsymbol{W}^{H/2}\boldsymbol{Y}\boldsymbol{Y}^H\boldsymbol{W}^{1/2}\right) \\
&= \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{V}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)^{1/2}\boldsymbol{V}_x^H\boldsymbol{\Phi}^{-\frac{1}{2}}\mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}^H)\boldsymbol{\Phi}^{-\frac{1}{2}}\boldsymbol{V}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)^{1/2}\boldsymbol{V}_x^H\boldsymbol{\Phi}^{\frac{1}{2}} \\
&= \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{V}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)^{1/2}\boldsymbol{\Lambda}_x\left(\mathbf{I} - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\right)^{1/2}\boldsymbol{V}_x^H\boldsymbol{\Phi}^{\frac{1}{2}} \\
&= \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{V}_x\left(\boldsymbol{\Lambda}_x - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_\eta)}{M-\hat{N}}\boldsymbol{I}\right)\boldsymbol{V}_x^H\boldsymbol{\Phi}^{\frac{1}{2}} = \boldsymbol{\Gamma}_x.
\end{aligned} \tag{4.19}$$

Here, $\mathcal{E}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^H\right) = \boldsymbol{\Gamma}_x$ as expected.

4.2 Known normalised spectral spatial matrix $\boldsymbol{\Psi}$ with an inhomogeneous noise field

This section first considers the same case as in the previous section with $N < M$, but we no longer suppose that all noise PSDs Γ_{η_i} are the same for each microphone. Then, the following system follows (refer to section 2.3.1 for definitions):

$$\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_x + \sqrt{\mathbf{D}}\boldsymbol{\Phi}\sqrt{\mathbf{D}}. \tag{4.20}$$

We can make a distinction between two cases, the first is when the normalised spatial spectral(or correlation matrix) $\boldsymbol{\Psi}$ is known (section 2.3.1), and the second is when we know the coherence matrix $\boldsymbol{\Phi}$. The last case is much more difficult to solve. The equivalent of inhomogeneous noise could also arise in the special situation when an M^{th} source with vector transfer function $\boldsymbol{H}_v$ survenes where there was previously $M-1$ sources in an homogeneous diffuse noise field, in such a way that the number of sources becomes equal to the number of microphones ($N = M$) and considering that the new source is part of the noise field:

$$\begin{aligned}\boldsymbol{\Gamma}_y &= \boldsymbol{\Gamma}_x + \boldsymbol{H}_v\boldsymbol{H}_v^H\Gamma_v + \Gamma_\eta\boldsymbol{\Phi} \\ &= \boldsymbol{\Gamma}_x + \sqrt{\mathbf{D}_{\text{mod}}}\boldsymbol{\Phi}_{\text{mod}}\sqrt{\mathbf{D}_{\text{mod}}},\end{aligned}$$

where

$$\mathbf{D}_{\text{mod}} = diag\left(\Gamma_\eta + \left|H_{v,1}\right|^2\Gamma_v, \dots, \Gamma_\eta + \left|H_{v,N}\right|^2\Gamma_v\right),$$

$$\boldsymbol{\Phi}_{\text{mod}} = \sqrt{\mathbf{D}_{\text{mod}}^{-1}}\left(\boldsymbol{H}_v\boldsymbol{H}_v^H\Gamma_v + \Gamma_\eta\boldsymbol{\Phi}\right)\sqrt{\mathbf{D}_{\text{mod}}^{-1}}. \tag{4.21}$$

However, this special case will not be developed further, since its form is not easily mathematically tractable. It was only mentioned here for the sake of completeness.

If $\boldsymbol{\Psi}[r,k]$ is known and using the relation of Eq.(2.74) ,then Eq.(4.20) can be rewritten as:

$$\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_x + \Gamma_v\boldsymbol{\Psi} \tag{4.22}$$

We notice immediately that this equation is in the same form as Eq.(2.112). We can thus use the same arguments of section 0 to find Γ_v. So in this case, we want to find the minimal value of Γ_v that solves the following characteristic polynomial,

$$\det\left(\boldsymbol{\Gamma}_y - \Gamma_v\boldsymbol{\Psi}\right) = 0. \tag{4.23}$$

All the algorithms are the same as in section 4.1 except that we use $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$**.** In fact, when the noise field is spherical or cylindrical $\boldsymbol{\Psi} = \boldsymbol{\Phi}$ and $\Gamma_v = \Gamma_\eta$, for example in an acoustic free field. An example of a situation when we would need to use $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$ in free field conditions is when the microphones do not have the same directivity gain. The estimate of the noise PSD matrix will be using the notation for the i$^{\text{th}}$ generalized eigenvalue $\lambda_i\left(\boldsymbol{\Gamma}_y, \boldsymbol{\Psi}\right)$:

$$\lambda_{min}\left(\boldsymbol{\Gamma}_y, \boldsymbol{\Psi}\right)\boldsymbol{\Psi} = \widehat{\boldsymbol{\Gamma}}_\eta^{(1)}. \tag{4.24}$$

An important detail is that if we use a positively scaled version of the matrix $\boldsymbol{\Psi}$ it will not change the result because

$$\lambda_{min}(\boldsymbol{\Gamma}_y, \alpha\boldsymbol{\Psi}) = \frac{1}{\alpha}\lambda_{min}(\boldsymbol{\Gamma}_y, \boldsymbol{\Psi}), \tag{4.25}$$

and

$$\lambda_{min}(\boldsymbol{\Gamma}_y, \alpha\boldsymbol{\Psi})\alpha\boldsymbol{\Psi} = \frac{1}{\alpha}\lambda_{min}(\boldsymbol{\Gamma}_y, \boldsymbol{\Psi})\alpha\boldsymbol{\Psi} = \lambda_{min}(\boldsymbol{\Gamma}_y, \boldsymbol{\Psi})\boldsymbol{\Psi} = \widehat{\boldsymbol{\Gamma}}_\eta^{(1)}. \tag{4.26}$$

This permits us to use any scaled correlation matrix to estimate the noise PSD matrix. The situation may occur whenever we estimate $\boldsymbol{\Psi}$ using identically scaled versions of the HRTFs. The same holds true for the homogeneous case, but since $\boldsymbol{\Phi}$ is by definition 1 at its diagonals, it generally isn't multiplied by an unwanted scalar.

4.2.1 First estimate: Approximating the noise PSD matrix

Eq.(4.24) is an algebraic solution. Using the maximum likelihood estimate in Eq.(2.143) of the noise PSD at the radiating surface we get a first estimate of the noise PSD matrix:

$$\widehat{\boldsymbol{\Gamma}}_\eta^{(1)} = \frac{tr(\boldsymbol{\Lambda}_v)}{M - \widehat{N}}\boldsymbol{\Psi}. \tag{4.27}$$

4.2.2 Second estimate: Noise subspace update

The estimate can be further improved using the similar update of the noise subspace on $\widehat{\boldsymbol{\Gamma}}_\eta^{(1)}$ using projection matrices defined as

$$\begin{aligned}\boldsymbol{P}_v &= \boldsymbol{V}_v(\boldsymbol{V}_v^H\boldsymbol{V}_v)^{-1}\boldsymbol{V}_v^H = \boldsymbol{P}_{x^\perp} = \boldsymbol{I} - \boldsymbol{P}_x,\\ \boldsymbol{P}_x &= \boldsymbol{V}_x(\boldsymbol{V}_x^H\boldsymbol{V}_x)^{-1}\boldsymbol{V}_x^H = \boldsymbol{P}_{v^\perp} = \boldsymbol{I} - \boldsymbol{P}_v,\end{aligned} \tag{4.28}$$

where

$$\boldsymbol{\Psi}^{-1/2}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Psi}^{-1/2} = [\boldsymbol{V}_x\ \boldsymbol{V}_v]\begin{bmatrix}\boldsymbol{\Lambda}_x & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Lambda}_v\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x^H\\ \boldsymbol{V}_v^H\end{bmatrix}.$$

The estimate of $\boldsymbol{\Gamma}_\eta$ can be further improved by updating the instantaneous noise subspace:

$$\widehat{\boldsymbol{\Gamma}}_\eta^{(2)} = \boldsymbol{\Psi}^{\frac{1}{2}}\left(\boldsymbol{P}_v\boldsymbol{\Psi}^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_y\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{P}_v + \boldsymbol{P}_x\boldsymbol{\Psi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Gamma}}_\eta^{(1)}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{P}_x\right)\boldsymbol{\Psi}^{\frac{1}{2}}. \tag{4.29}$$

4.2.3 Third estimate: Correlation matrix update

Because of the scale invariance property in Eq.(3.32), we can use the following matrix as an estimate of an instantaneous correlation matrix:

$$\widehat{\boldsymbol{\Psi}}'[r,k] = \frac{M\hat{\boldsymbol{\Gamma}}_\eta[r,k]}{tr\left(\hat{\boldsymbol{\Gamma}}_\eta[r,k]\right)}. \tag{4.30}$$

By accumulating the changes with a first order recursion, we can estimate the correlation matrix with

$$\widehat{\boldsymbol{\Psi}}_r = \mu\widehat{\boldsymbol{\Psi}}_{r-1} + (1-\mu)\widehat{\boldsymbol{\Psi}}'[r,k], \tag{4.31}$$

where

$$\widehat{\boldsymbol{\Psi}}[r,k] = \widehat{\boldsymbol{\Psi}}_r,$$

with initial condition

$$\widehat{\boldsymbol{\Psi}}[-1,k] = \boldsymbol{\Psi}.$$

Again, the smoothing parameter μ needs to be close to one and any reasonable initial condition could be used, e.g. $\widehat{\boldsymbol{\Psi}}[-1,k] = \boldsymbol{\Phi},\boldsymbol{\Psi},\mathbf{I}$. Estimating $\widehat{\boldsymbol{\Psi}}$ in this fashion might give much worse results than using the coherence matrix ($\widehat{\boldsymbol{\Phi}}$) estimation since all values in $\widehat{\boldsymbol{\Psi}}$ vary whereas only the off diagonal terms in $\widehat{\boldsymbol{\Phi}}$ do. For example, when the diagonals of $\widehat{\boldsymbol{\Psi}}$ are approximately equal and unity, it would be better to estimate $\widehat{\boldsymbol{\Phi}}$ instead of $\widehat{\boldsymbol{\Psi}}$ since the diagonals of the latter might not be equal and contribute to propagating errors in $\widehat{\boldsymbol{\Psi}}$ along the time frames. In other words, estimating $\widehat{\boldsymbol{\Phi}}$ allows less degrees of freedom for the errors to propagate into.

The final adaptive estimate is

$$\widehat{\boldsymbol{\Gamma}}_{\eta}^{(3)} = \widehat{\boldsymbol{\Psi}}_{r-1}^{\frac{1}{2}}\left(\boldsymbol{P}_{\eta}\widehat{\boldsymbol{\Psi}}_{r-1}^{-\frac{1}{2}}\widehat{\boldsymbol{\Gamma}}_{y}[r,k]\widehat{\boldsymbol{\Psi}}_{r-1}^{-\frac{1}{2}}\boldsymbol{P}_{\eta} + \frac{\mathrm{tr}(\boldsymbol{\Lambda}_{\eta})}{M-\widehat{N}}\boldsymbol{P}_{x}\right)\widehat{\boldsymbol{\Psi}}_{r-1}^{\frac{1}{2}}. \quad \textbf{4.32}$$

If α is too small then estimating the signal subspace becomes less reliable and so, there is a possibility that $\widehat{\boldsymbol{\Phi}}$ might actually diverge from the environments true coherence matrix. To obtain a good $\widehat{\boldsymbol{\Psi}}$ we need a good $\widehat{N}$ but a good estimate of $\widehat{N}$ needs a reliable $\widehat{\boldsymbol{\Psi}}$. This means that the above algorithm(third estimate) might never converge. As mentioned before, a way around this would be to estimate N using methods that don't rely on $\widehat{\boldsymbol{\Psi}}$. Such an estimator of N is analogous to voice activity detection (VAD).

Summarizing the proposed algorithms of section 4.2

Except for the use of $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$ and the instantaneous update of $\boldsymbol{\Psi}$, the algorithms are identical to the ones presented in section 4.1.

Step 1 Compute $\widehat{N}$ with the AIC or set $\widehat{N}$ at 1.

Step 2 Compute $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)} = \frac{\mathrm{tr}(\boldsymbol{\Lambda}_{\nu})}{M-\widehat{N}}\boldsymbol{\Psi}$ with eigendecomposition described in Eq.(4.28).

Step 3 Update the noise subspace of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$with $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)} = \boldsymbol{\Psi}^{\frac{1}{2}}\left(\frac{\mathrm{tr}(\boldsymbol{\Lambda}_{\nu})}{M-\widehat{N}}\boldsymbol{P}_{x} + \boldsymbol{V}_{\nu}\boldsymbol{\Lambda}_{\nu}\boldsymbol{V}_{\nu}^{H}\right)\boldsymbol{\Psi}^{\frac{1}{2}}$.

Step 4 Update the correlation matrix with Eq.(4.30) and Eq.(4.31).

Table 6: Correlated matrix based noise PSD estimation methods summary.

Steps 1 and 2 corresponds to the first estimate. The second and third estimates correspond to steps 3 and 4 respectively.

4.3 Known coherence matrix $\mathbf{\Phi}$ with an inhomogeneous noise field and known source transfer function

4.3.1 Algebraic solution

The situation where we only know the coherence matrix (potentially different from the correlation matrix $\mathbf{\Psi}$) to model a inhomogeneous noise field gives rise to a very different set of solutions. But it is presented as being more of a theoretical solution rather than a practical one since if we can model $\mathbf{\Phi}$, we can probably model $\mathbf{\Psi}$ and hence we fall in the previous section 4.2. To solve Eq.(4.20) we can isolate $\mathbf{\Gamma}_x$ and take the determinant on both sides to get the following equation

$$\begin{aligned} \det\left(\boldsymbol{\Gamma}_y - \sqrt{\mathbf{D}}\mathbf{\Phi}\sqrt{\mathbf{D}}\right) &= \det(\mathbf{\Gamma}_x), \\ \det\left(\boldsymbol{\Gamma}_y - \sqrt{\mathbf{D}}\mathbf{\Phi}\sqrt{\mathbf{D}}\right) &= 0 \text{ (since } \mathbf{\Gamma}_x \text{ is rank defective).} \end{aligned} \tag{4.33}$$

The equation is a function of multiple variables and it has an infinity of solutions. It can be seen from Eq.(4.20) that it is an underdetermined system. However for the case of a single source and if the transfer functions of the source are known, the system of equations can be solved. Let's present this case to see how the equations can quickly become complicated. Assuming that the transfer functions of the source is known, that they are approximately equal to 1 (e.g., special case of near-frontal source), with two microphones then the system of equations to solve is

$$\begin{bmatrix} \Gamma_{y1} & \Gamma_{y12} \\ \Gamma_{y12}^* & \Gamma_{y2} \end{bmatrix} - \begin{bmatrix} \Gamma_{\eta_1} & \sqrt{\Gamma_{\eta_1}\Gamma_{\eta_2}}\Phi \\ \sqrt{\Gamma_{\eta_1}\Gamma_{\eta_2}}\Phi & \Gamma_{\eta_2} \end{bmatrix} = \Gamma_s \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{4.34}$$

Looking at the determinants, we have

$$\det\left(\begin{bmatrix} \Gamma_{y1} & \Gamma_{y12} \\ \Gamma_{y12}^* & \Gamma_{y2} \end{bmatrix} - \begin{bmatrix} \Gamma_{\eta_1} & \sqrt{\Gamma_{\eta_1}\Gamma_{\eta_2}}\Phi \\ \sqrt{\Gamma_{\eta_1}\Gamma_{\eta_2}}\Phi & \Gamma_{\eta_2} \end{bmatrix} \right) = 0. \tag{4.35}$$

We know the transfer functions of the vector transfer function, so if we substitute the values,

$$\begin{aligned} \Gamma_{\eta_1} &= \Gamma_{y1} - \Gamma_s, \\ \Gamma_{\eta_2} &= \Gamma_{y2} - \Gamma_s. \end{aligned} \tag{4.36}$$

Then

$$\begin{aligned} \det\begin{pmatrix} \Gamma_s & \Gamma_{y12} - \sqrt{(\Gamma_{y1} - \Gamma_s)(\Gamma_{y2} - \Gamma_s)}\Phi \\ \Gamma_{y12}^* - \sqrt{(\Gamma_{y1} - \Gamma_s)(\Gamma_{y2} - \Gamma_s)}\Phi & \Gamma_s \end{pmatrix} \\ = \Gamma_s^2 - |\Gamma_{y12}|^2 - (\Gamma_{y1} - \Gamma_s)(\Gamma_{y2} - \Gamma_s)\Phi^2 \\ + 2\mathrm{Re}(\Gamma_{y12})\sqrt{(\Gamma_{y1} - \Gamma_s)(\Gamma_{y2} - \Gamma_s)}\Phi = 0. \end{aligned} \tag{4.37}$$

We can modify the nonlinear equation to a quartic one. If we isolate the square root term and raise it to the power of two we get the desired polynomial. The problem also becomes the one of finding the signal power density Γ_s instead of the noise power densities. Finding the roots of the quartic polynomial yields two negative numbers and two positive numbers. Since the roots are real, they can be ordered from largest to smallest. The negative numbers are obviously rejected since power is a positive quantity. The correct solution is the minimum of the two positive numbers. It is interesting to note that actually the two smallest roots in magnitude are solutions to the original equation. However, in practice the polynomial will yield complex roots, and so no solution is reliable.

4.3.2 Applying single-channel algorithms to the multi-channel case

When we look at Eq.(2.29) we can't help but notice that if we do have at our disposal $\mathbf{\Phi}$ then we only are left to estimate

$$\mathbf{D} = Diag\left(\Gamma_{\eta_1}, \Gamma_{\eta_2}, \dots, \Gamma_{\eta_M}\right). \tag{4.38}$$

The idea of estimating the above matrix (diagonal elements) using single-channel noise estimating algorithms [21, 36-38] was mentioned in [4]. When this is done we are left with the noise PSD matrix estimate of

$$\hat{\mathbf{\Gamma}}_{\eta}^{(SC)} = \sqrt{\mathbf{D}}\mathbf{\Phi}\sqrt{\mathbf{D}}. \tag{4.39}$$

In this thesis, only the diagonal terms of $\mathbf{D}$ will be estimated. We will not evaluate the performance measures of $\hat{\mathbf{\Gamma}}_{\eta}$ using all the matrices terms for example with Eqs.(2.106,2.108,2.110) , only Eq.(2.100) will be used as a performance measure.

4.4 Estimation of single source PSD in a mixture of several sources

For this scenario we first make the assumption that there are $N = M$ sources in total (same as the number of microphones), this restrictive assumption will be dropped later. We seek to estimate the PSD of a source for which we know the transfer functions i.e., $\boldsymbol{H}$ is known. For notational convenience, we denote this PSD by Γ_{ν}. Note that this source can be a target or an interferer. It is also no longer assumed that there is an isotropic acoustic noise field, in fact we assume here that there is no such field. Then, the following system follows:

$$\mathbf{\Gamma}_y = \mathbf{\Gamma}_x + \boldsymbol{H}\boldsymbol{H}^H\Gamma_{\nu}. \tag{4.40}$$

This system of equations has the same form as in Eq.(4.22) except with $\mathbf{\Psi} = \boldsymbol{H}\boldsymbol{H}^H$,and from Eq.(2.24), $\mathbf{\Gamma}_{\eta} = \boldsymbol{H}\boldsymbol{H}^H\Gamma_{\nu}$. So

$$\det\left(\boldsymbol{\Gamma}_y - \boldsymbol{H}\boldsymbol{H}^H\Gamma_{\nu}\right) = \det\left(\boldsymbol{\Gamma}_y\right) - \Gamma_{\nu}\boldsymbol{H}^H\boldsymbol{\Gamma}_y^A\boldsymbol{H} = 0, \tag{4.41}$$

where $\boldsymbol{B}^A$ denotes the adjoint of a matrix $\boldsymbol{B}$. The solution is then

$$\Gamma_{\nu} = \left(\boldsymbol{H}^H\boldsymbol{\Gamma}_y^{-1}\boldsymbol{H}\right)^{-1}. \tag{4.42}$$

We note that this is identical to a minimum variance distortionless response (MVDR) beamformer gain[22]. Since in practice $\mathbf{\Gamma}_y$ will always be invertible[13] for a number of samples L much bigger than M, this means that we can use Eq.(4.42) as an approximation of the desired noise PSD for any number of sources whenever the number of terms to approximate $\mathbf{\Gamma}_y$ is sufficiently large (i.e., it is not required that the total number of sources be equal to M). On the other side, when effectively $\mathbf{\Gamma}_y$ is ill-conditioned or not invertible, we would need to use some regularization analog to diagonalization:

$$\Gamma_v = \left(\boldsymbol{H}^H\left(\mathbf{\Gamma}_y + \delta\mathbf{\Psi}\right)^{-1}\boldsymbol{H}\right)^{-1}, \tag{4.43}$$

with δ sufficiently small. The reason why $\mathbf{\Psi}$ would be used instead of $\boldsymbol{I}$ is because we need to simulate a very small power additive diffuse noise associated to the acoustic environment. But $\boldsymbol{I}$ could also be used say for its inherent simplicity which would correspond to classical diagonalization.

This case of estimating a single source PSD in a mixture of several sources will not be simulated in this thesis since it is much related to basic beamforming which is already a well-developed area of research. For the same reason, other estimators that presume that we know multiple directional source vectors(or transfer functions) whenever there is no diffuse acoustic field will also be omitted.

[13] If $\mathbf{\Gamma}_y$ is invertible it doesn't necessarily means that it is well conditioned. The only time where $\mathbf{\Gamma}_y$ is not invertible is when the number of terms used to estimate $\mathbf{\Gamma}_y$ is under the number of microphones M. When the number of samples is of the order of M, it can happen that $\mathbf{\Gamma}_y$ is ill-conditioned.

Chapter 5 Simulation results (★)

5.1 Simulation settings

The simulations are done using the Oldenburg University database [11] and the TIMIT database. The sentences taken from the TIMIT database are used to simulate three speakers and all recordings have a 16 bits resolution and a 16 kHz sampling rate. Sentences are concatenated to sum to approximately 15 seconds. The data from [11] contain HRIRs and ambient noise measured in different acoustic environments. All the data has a resolution of 32 bits and a sampling rate of 48 kHz. The TIMIT sentences consequently are resampled from 16 kHz to 48 kHz using MATLAB's "*resample*" function. To reproduce the sources at the different microphones, the HRIRs from the Oldenburg database are convolved[14] with TIMIT sentences to simulate speakers in a three dimensional acoustic environment. The HRIR measurements were recorded with microphones arranged as in the following figure:

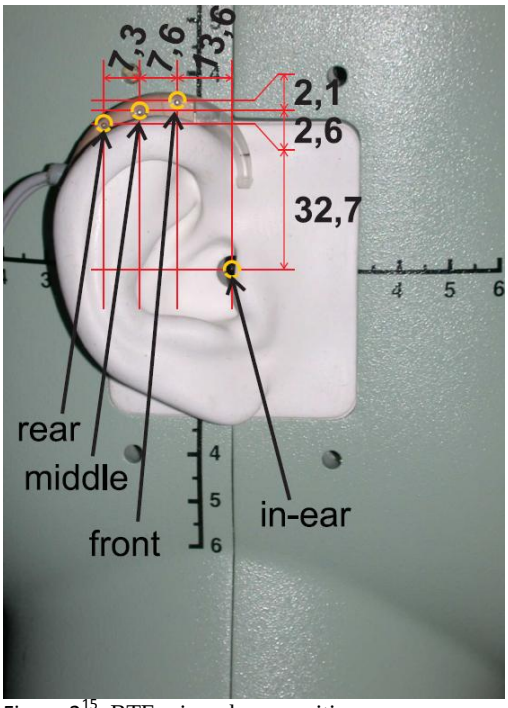

Figure 2[15]: BTE microphone positions.

Eight microphones were mounted on a dummy head with 4 microphones per ear. The top three microphones are from a behind-the-ear (BTE) hearing aid, while the 4th microphone is located in the ear canal.

In this chapter simulations are done using the cafeteria environment and anechoic chamber environment. The first, second and third simulated sources come respectively from the directions "A", "B" and "D" from the cafeteria setting presented in the following figure:

[14] It is more effective to implement a convolution by a multiplication in the Fourier domain.

[15] Image courtesy of Kayser et al.[11], available online.

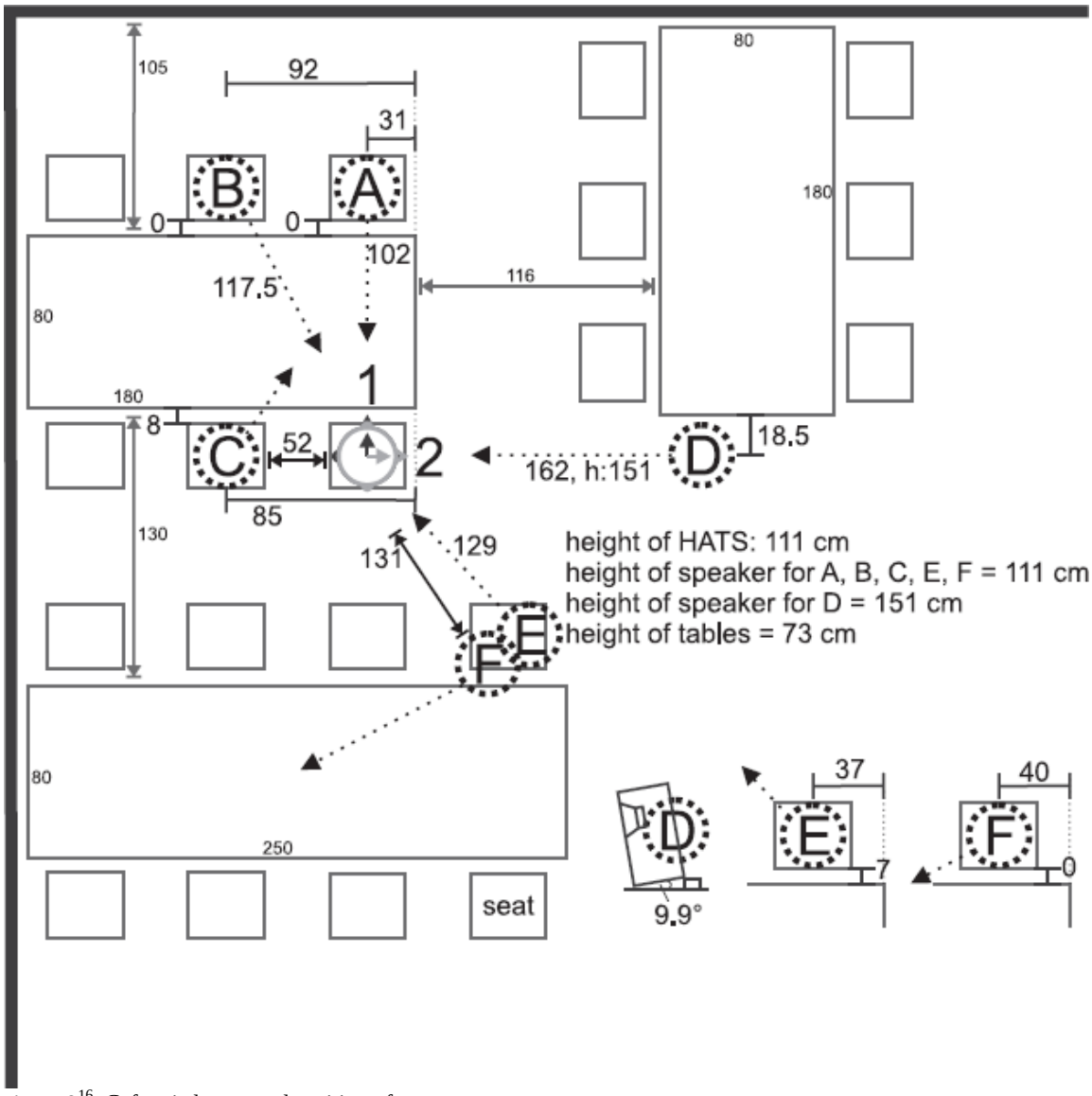

Figure 3[16]: Cafeteria layout and position of sources.

The sources from the anechoic chamber are simulated to be at the angles prescribed by the following table:

Source	Azimuth ϕ	Elevation θ
1	0°	0°
2	35°	0°
3	-60°	0°

Table 7: Sources angles of arrival in the anechoic environment.

[16] Image courtesy of Kayser et al.[11].

The coherences and correlations were computed using the HRIRs measured in the anechoic chamber, i.e. $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are computed using the Riemann integral approximation in Eq.(2.77) and Eq.(2.78) respectively.

Simulations are performed for the number of sources $N \in [0,3]$. The case with no source is the reference values for the comparison which is done using the performance metric of Eq.(2.103) if the homogeneous assumption is made, and the performance metric of Eq.(2.105) if the noise field is inhomogeneous.

The "best" algorithm is the one that gives the lowest log-error. The recursive parameter α is set to 0.7 to compute the matrix periodogram using the recursive average technique prescribed by Eq.(2.139). The window length is chosen to be 20msec because we want to have a good time resolution. The analysis segment is chosen to be relatively short to pick fast changes of the noise PSD. A Hamming window is used in computing the analysis segments. The simulated input SNR defined in Eq.(2.99) will be varied from 0 to 15dB by increments of 5dB. Note that our definition of source here includes all the directional sources, possibly including both targets and interferers. For the noisy signal to have the desired iSNR, the simulated noisy signal $\boldsymbol{y}[n]$ has the form:

$$\boldsymbol{y}[n] = \sqrt{a}\boldsymbol{x}[n] + \boldsymbol{\eta}[n],$$

where

$$a = 10^{\frac{iSNR}{10}} \cdot \frac{\sum_{n,i} \eta_i^2[n]}{\sum_{n,i} x_i^2[n]}. \qquad \textbf{5.1}$$

Except for the SNR adjustment, no scaling was done for the databases used i.e., the sentences, HRTFs and the cafeteria noise were taken as is[17]. An identification code for the channels taken from [11] will be used throughout the rest of the present work, and is given by the following table:

Channel(cafeteria)	Channel(anechoic)	Channel Id
Left in-ear	Left in-ear	1
Right in-ear	Right in-ear	2
Left rear BTE	Left front BTE	3
Right rear BTE	Right front BTE	4
Left middle BTE	Left middle BTE	5
Right middle BTE	Right middle BTE	6
Left front BTE	Left rear BTE	7
Right front BTE	Right rear BTE	8

Table 8: Channel identification codes

[17] In[8] the signal scaling is different from this thesis. Our scaling is smaller and will result in much smaller PSDs. Double floating point arithmetic is used and, the problem is well conditioned(since $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are well conditioned).

Four cases are considered and are shown in the following table:

	Channel Id set ($\mathcal{M}$)
Binaural, $M = 2$	{1,2}
Monaural, $M = 2$	{3,7}
Monaural, $M = 3$	{3,5,7}
Binaural, $M = 4$	{3,5,7,8}

Table 9: Channel sets identification codes

For example, if $\mathcal{M} = \{3,5,7\}$ this would correspond to the scenario using all the left side BTE hearing aid microphones.

Here is a list of parameters for different scenarios with $N \geq 1$:

$$Environment: \mathcal{En} = \{"Anechoic", "Cafeteria"\}$$
Number of Sources : $N \in [1,2,3]$
$$iSNR \in [0,5,10,15]dB$$
Microphone settings: $\mathcal{M} = \{\{1,2\},\{3,7\},\{3,5,7\},\{3,5,7,8\}\}$
Matrix used in the algorithm: $\mathcal{C} = \{\boldsymbol{\Psi}, \boldsymbol{\Phi}\}$

There is a total of $2 \cdot 3 \cdot 4 \cdot 4 \cdot 2 = 192$ different scenarios to simulate.

5.2 Single-channel algorithms

The single channel schemes chosen for the simulation are Martin's Minimal Statistics (MS), as well as the algorithms from Doblinger and Hendriks et al. (MMSE) [21, 37, 38]. The algorithms are used to estimate the auto-noise PSD for each channel. In this section the noise PSD of all the channels is estimated. Each case is done in the anechoic and the cafeteria environments and the SNR is varied. The log-error performance measures taking each channel independently (Eq.(2.100)) are presented in the following tables for the number of sources $N \in [1,3]$. The log-error values below are averaged in time and in frequencies up to 10KHz. Since the tables are large, colors will be used to indicate the lowest cost with respect to the number of sources and the SNR. Blue will refer to the MMSE algorithm and red will refer to the MS one.

In the anechoic room simulation, estimators seem to score better in general, as we can see in the table below where the log-error is averaged in frequencies up to 10 KHz.

	MS algorithm				**Doblinger's algorithm**				**MMSE based algorithm**			
	SNR				**SNR**				**SNR**			
N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
1	2,7	2,6	2,6	2,6	3,2	3,1	3,2	3,6	**1,8**	**1,9**	**2,0**	**2,2**
2	2,7	2,6	2,7	3,0	3,2	3,1	3,3	3,9	**1,9**	**2,1**	**2,4**	**2,9**
3	2,7	2,7	2,8	3,3	3,2	3,1	3,4	4,1	**2,0**	**2,3**	**2,7**	**3,2**

Table 10: Log-error in anechoic environment for varying sources, SNRs and noise estimation algorithms.

It is clear from the table that the MMSE based algorithm is superior on average across the different scenarios. The log-error cost in the frequency and time domain is shown next for the specific case of N=3 and SNR=15dB. We notice that although the MMSE based algorithm yields good average scores, it has a high variance in the cost along the frequency domain.

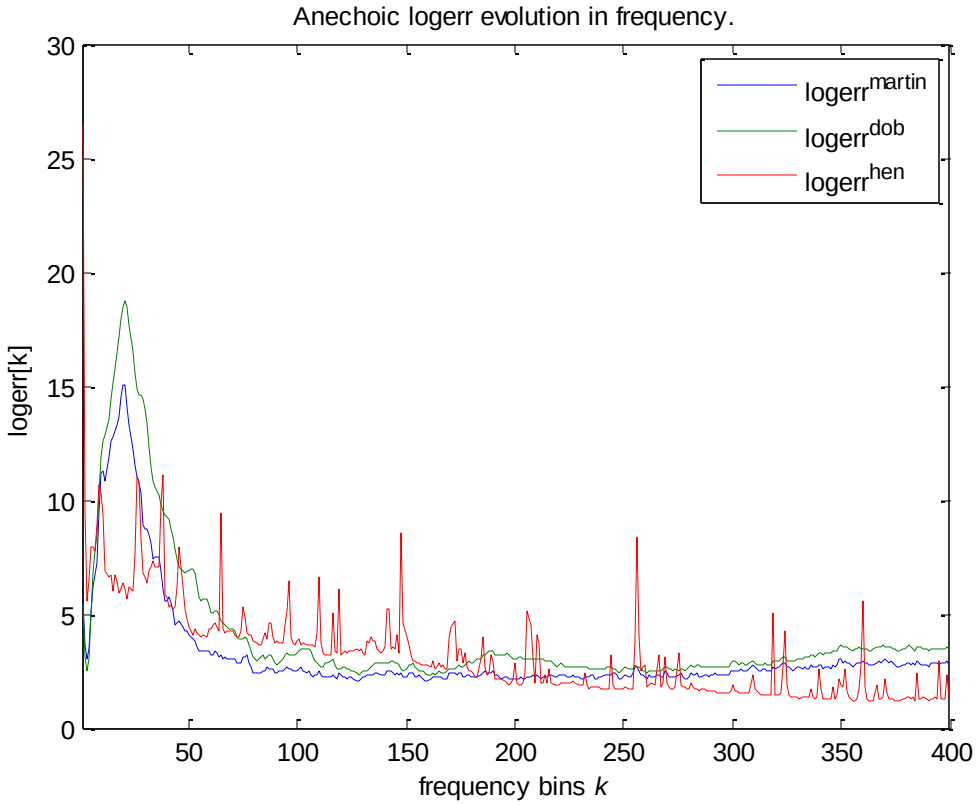

Figure 4: Anechoic log-error evolution in frequency with N=3,SNR=15dB.

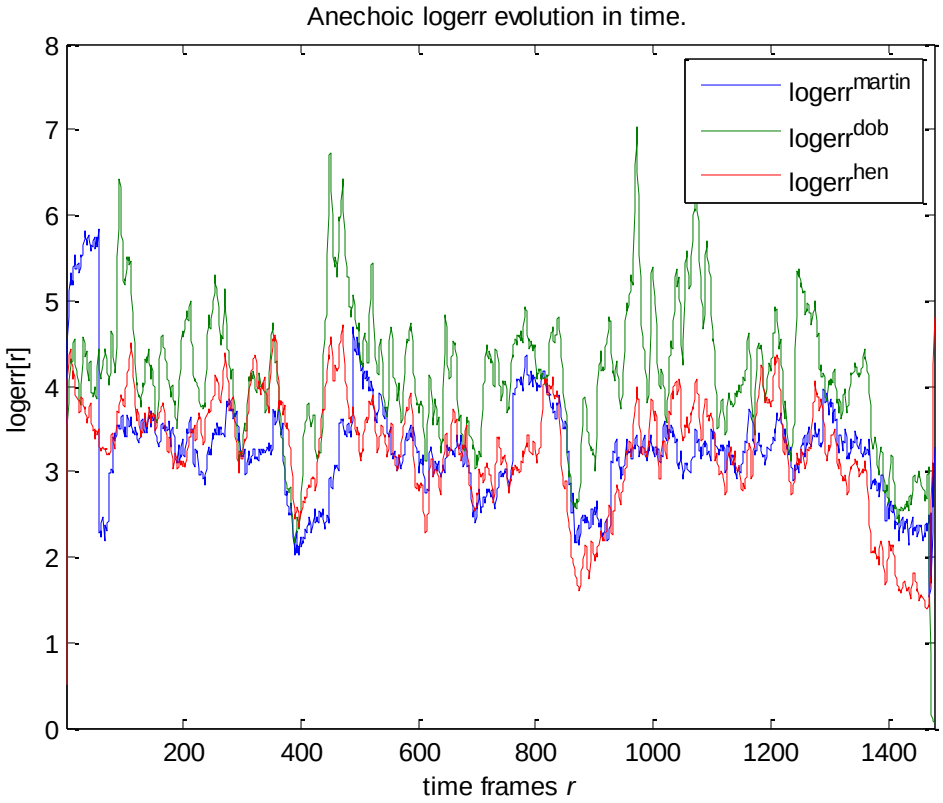

Figure 5: Anechoic log-error evolution in time with *N*=3,SNR=15dB.

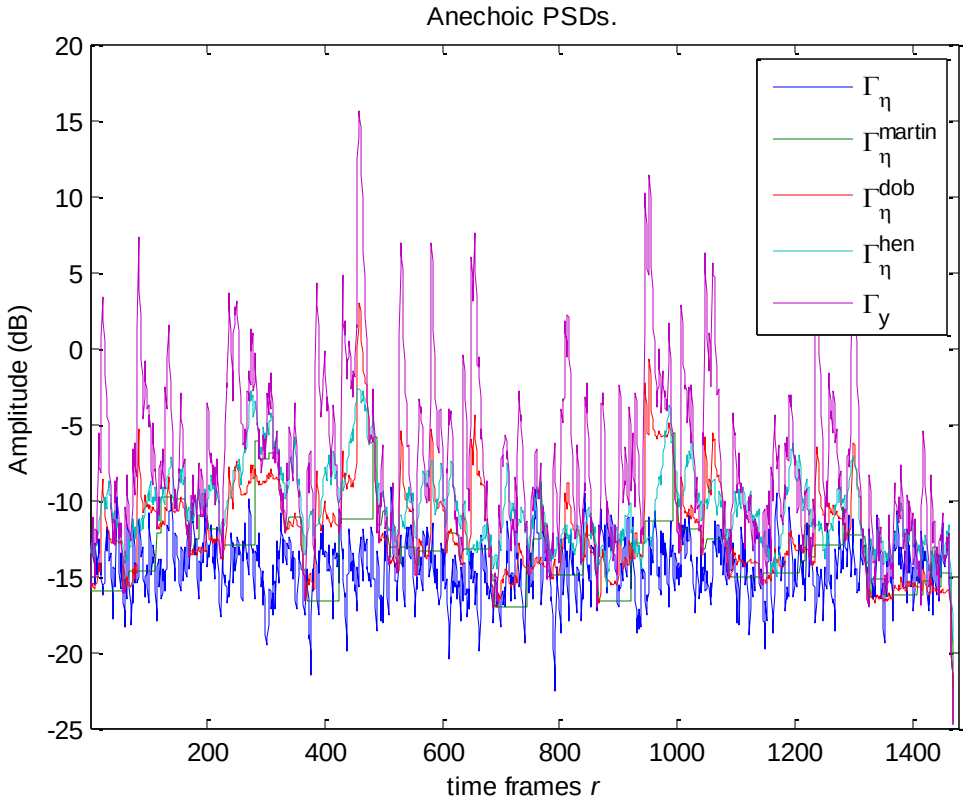

Figure 6: Anechoic PSDs at 2500 Hz with *N*=3,SNR=15dB, in reference to the left in-ear microphone.

For the same specific case of *N*=3 and SNR=15dB, in the above graphic the MMSE noise PSD (labeled with hen) is overestimated for frequency 2500Hz just as the PSD computed with Doblinger's method. So for this particular case, Martin's MS algorithm is better since in average it overestimates less the noise power level. This illustrates that each algorithm will have its own setups where it outperforms the others.

	MS algorithm				**Doblinger's algorithm**				**MMSE based algorithm**			
	SNR				**SNR**				**SNR**			
N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
1	4,1	4,0	3,8	3,9	4,2	4,1	4,3	4,9	**2,5**	**2,6**	**2,9**	**3,4**
2	4,0	3,8	3,8	**4,2**	4,0	3,9	4,2	5,3	**2,6**	**2,9**	**3,5**	4,4
3	3,9	3,7	3,8	**4,4**	4,0	3,8	4,2	5,3	**2,6**	**3,0**	**3,7**	4,8

Table 11: log-error in the cafeteria for varying sources and SNRs and noise estimation algorithms.

Considering now the results for the reverberant cafeteria environment, from the above table we notice that, averaged across the different setups, the MMSE based method is generally the best, followed by Martin's MS PSD estimator and Doblinger's algorithm. The MS algorithm becomes the best when speech is present most of the time, for example when *N*=3 at 15dB. We can observe this in the following two graphics that show the average log-error cost depending on frequency and time, respectively, again for this particular setup.

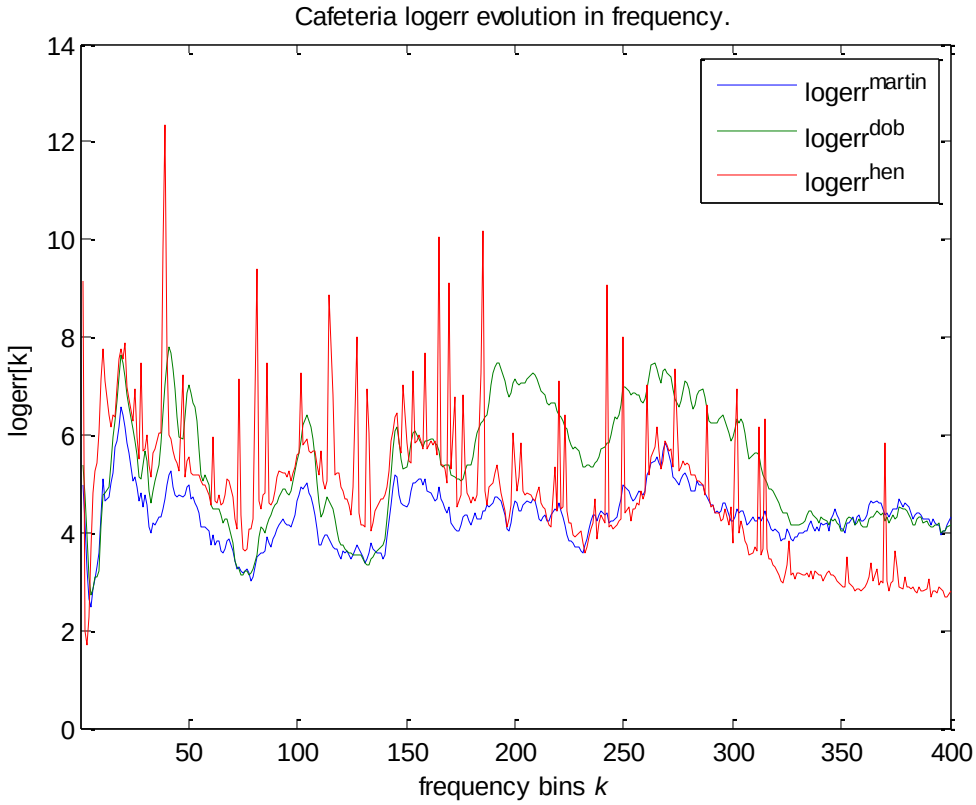

Figure 7: Cafeteria log-error evolution in frequency with N=3,SNR=15dB.

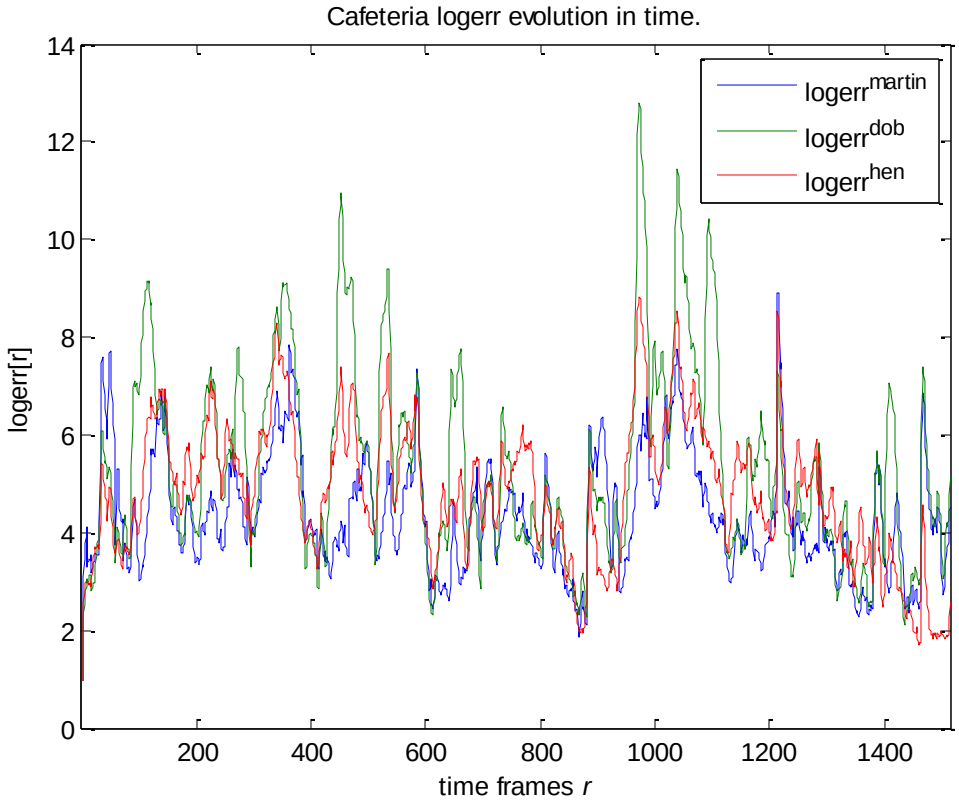

Figure 8: Cafeteria log. error evolution in time with N=3, SNR=15dB.

There is a strong correlation between the high cost values of the Doblinger algorithm (shown above) and speech activity. This can be validated in the next figure which shows the various PSDs estimates and the true PSDs, i.e., the noise and the noisy signal PSDs at the left in-ear microphone.

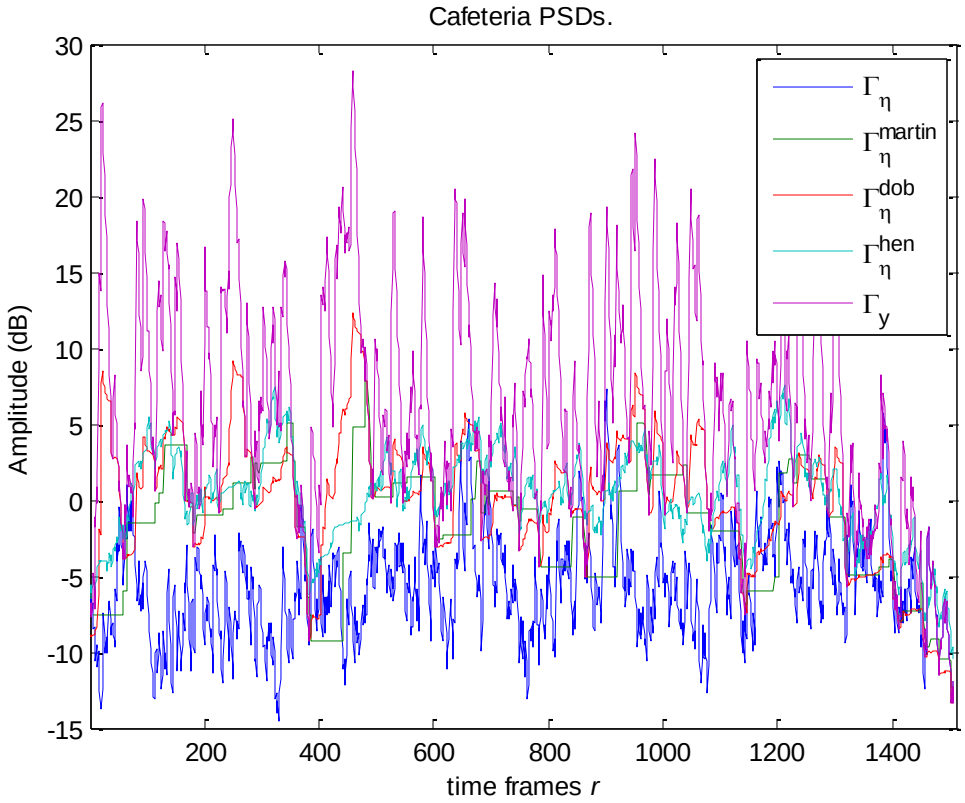

Figure 9: Cafeteria PSDs at 2500 Hz with *N*=3,SNR=15dB, in reference the left in-ear microphone.

5.3 Coherence based algorithms

The two-channel algorithms presented in section 3.2.1and 3.2.3(Robust dual-channel noise PSD estimation[1] and the PLDNE[2]) will not be simulated in this section because they require an *a priori* assumption on the directivity vector $\boldsymbol{H}_{M\times 1}$,while our proposed methods don't make this assumption. Recall that both of these algorithms are defined up to M=2 and use the hypothesis that the sources directivity are the unit vector $\mathbf{1}_{2\times 1}$. These restrictions don't suit all the different scenarios that we wish to consider and so the algorithms are not appropriate to simulate and to compare our results to. As for the other two-channel algorithm described in Chapter 3, i.e. the one in section 3.2.2, as previously mentioned it is equivalent conceptually to the algorithms presented in this thesis but it is limited to the subcase of $M = 2$ and $\widehat{N} = 1$, therefore it is implicitly simulated whenever $M = 2$ and $\widehat{N}$ is fixed at 1.

Four different scenarios are chosen by selecting appropriate microphones outputs.

1. 2-channel binaural setting using the in-ear microphones. (Channel set {1,2})
2. 2-channel monaural setting with microphones from the BTE on the left side of the head only (front and rear microphones, Channel set {3,7})
3. 3-channel monaural setting with microphones from the BTE on the left side of the head only. (Channel set {3,5,7})
4. 4-channel binaural setting using front and rear microphones from both the left and right side BTE hearing aid.(Channel set {3,4,7,8})

For the cases with $\mathcal{M} = \{\{3,7\},\{3,5,7\},\{3,5,7,8\}\}$, that is the cases where there is at least two closely spaced microphones, we can expect $\boldsymbol{\Psi}$ or $\boldsymbol{\Phi}$ to be ill-conditioned. Therefore the developed coherence based algorithms could possibly fail in such setups, for reasons previously explained.

5.3.1 Known time-invariant coherence or correlation matrix ($\boldsymbol{\Phi}$,$\boldsymbol{\Psi}$)

The elements of the coherence matrix and the correlation matrix are computed directly from the anechoic HRTFs using Eq.(2.77), assuming that the field is cylindrical. The right amount of truncation of time samples is used to make the coherence or correlation matrix have the same number of frequency bins as $\boldsymbol{\Gamma}_y$. We will compare the coherence based PSD estimation methods for both the assumption of homogeneous and inhomogeneous noise fields.

The following figure shows the correlation and coherence between both in-ear channels.

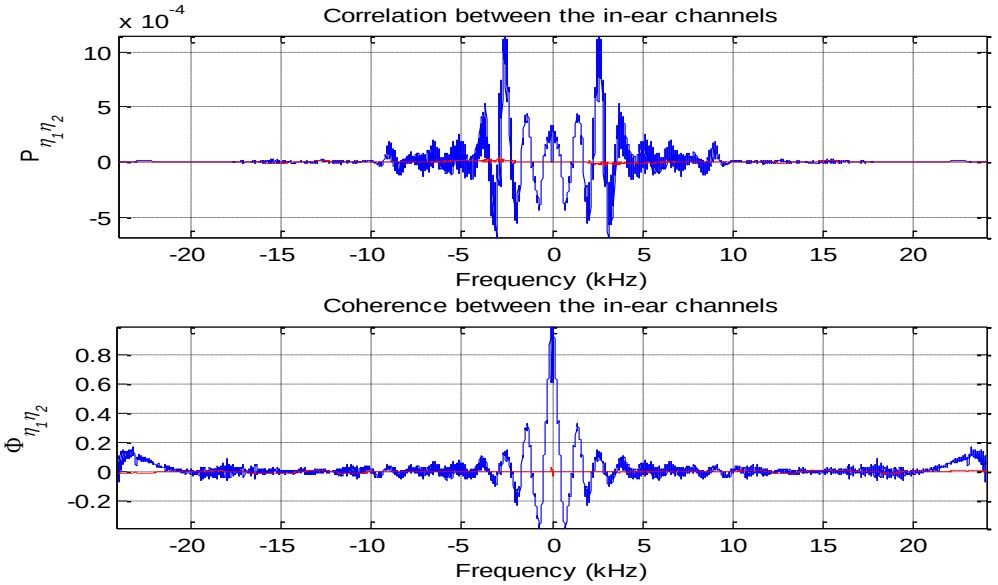

Figure 10: Coherence and correlation between in-ear channels.

In blue is the real part and in red is the imaginary part. We can see that the imaginary part is relatively negligible compared to the real part. It is so for the coherence and correlation when the pair of microphones have a symmetry e.g. left and right in-ear mic., left and right rear BTE mic., etc. When there is a slight asymmetry, e.g. when we compare left front BTE and right rear BTE, the imaginary part is no longer negligible, but it remains small as can be seen in the following figure:

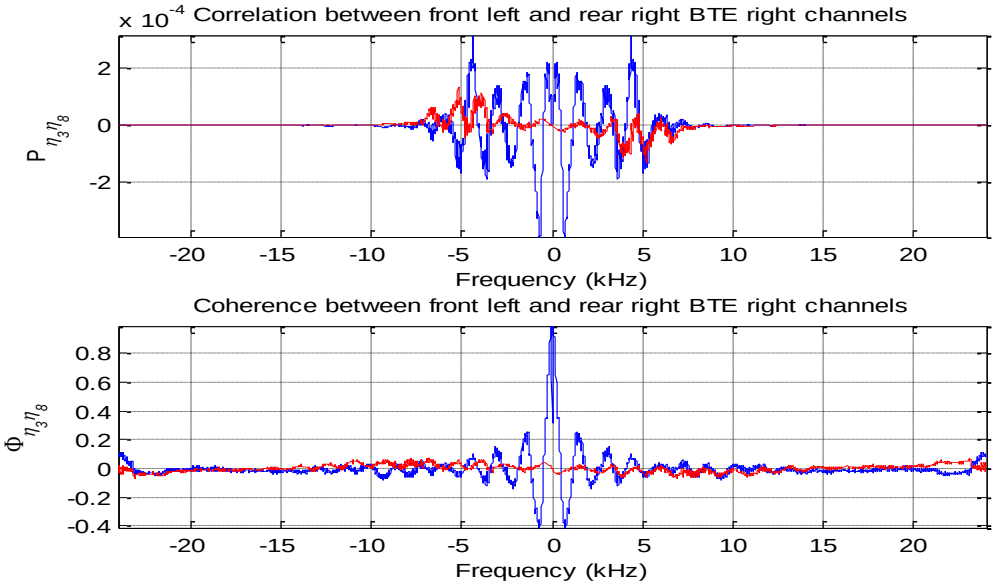

Figure 11: Coherence and correlation between front left and rear right BTE right channels.

When we compute the correlation and coherence of microphones on one side of the head only, the correlation will be high. It is also a very asymmetrical geometry and we can expect to see a non-negligible imaginary part, for example when the pair of microphones is the front and rear channel of the left side BTE unit. Also for this case, we can expect the main lobe of the real part to by wider because the microphone pair is close, for example in the following figure.

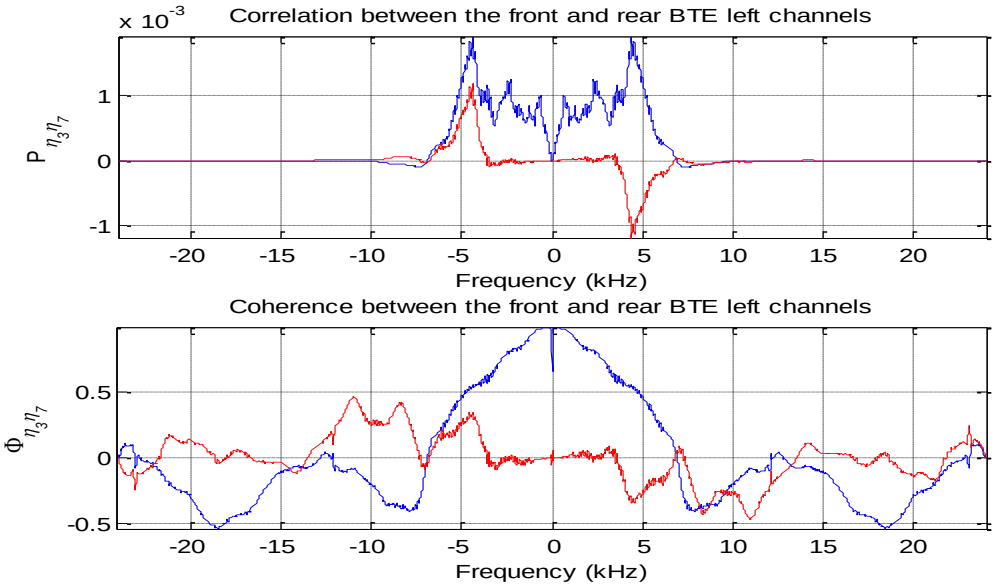

Figure 12: Coherence and correlation between front and rear BTE left channels

5.3.1.1 Tables of log-error distortion measures in anechoic environment

The first two tables in this section show the log-error distortion measure obtained with the proposed coherence-based methods for different scenarios, with a varying number of sources, and

with the homogeneous field assumption. Version 1 will be prescribed by Eq.(4.7) while Version 2 refers to Eq.(4.10).

		Version 1 SNR				Version 2 SNR			
	N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
Channels 1,2	**1**	2,5	2,4	2,3	2,3	2,5	2,4	2,3	2,3
	2	2,4	2,4	2,4	2,8	2,4	2,4	2,4	2,8
	3	2,5	2,5	2,8	3,4	2,5	2,5	2,8	3,4
Channels 3,7	**1**	2,6	2,5	2,4	2,3	2,6	2,5	2,4	2,3
	2	2,6	2,5	2,4	2,4	2,6	2,5	2,4	2,4
	3	2,6	2,5	2,5	2,7	2,6	2,5	2,5	2,7
Channels 3,5,7	**1**	**2,5**	**2,4**	**2,3**	**2,3**	3,0	2,8	2,6	2,4
	2	**2,5**	**2,5**	**2,5**	2,7	2,8	2,6	2,5	**2,5**
	3	**2,6**	**2,6**	2,8	3,3	2,9	2,7	**2,7**	**2,9**
Channels 3,4,7,8	**1**	2,2	2,1	2,1	2,1	**2,0**	**1,9**	**1,9**	**1,8**
	2	2,2	2,2	2,4	2,7	**2,0**	**2,0**	**2,0**	**2,3**
	3	2,3	2,4	2,6	3,2	**2,1**	**2,1**	**2,3**	**2,8**

Table 12: **Anechoic environment** log-error table with **fixed** $\widehat{\boldsymbol{N}} = \mathbf{1}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$.

For the 2-channel binaural case with fixed $\widehat{N} = 1$ our proposed coherence based method reduces to the previously published Kamkar-Parsi [3] algorithm of section 3.2.2 and for the anechoic environment in Table Table 12 it yields a higher cost than the 1-channel MMSE method, which is in general better than the MS algorithm (seen in Table Table 10). These results correspond to those reported in [20]. Our proposed estimate $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ can however provide the best performance for most conditions in the binaural 4-channel channel set $\{3,4,7,8\}$, even with a fixed $\widehat{N} = 1$ value. There are a few scenarios for the monaural 2 channel case set $\{3,4,7,8\}$ as well where the proposed method can produce the best results with a fixed $\widehat{N} = 1$ value.

The following table will now compare the two estimators using $\widehat{N}$ computed with the AIC:

		Version 1 SNR				Version 2 SNR			
	N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
Channels 1,2	**1**	1,6	1,7	1,7	1,8	**1,0**	**1,2**	**1,3**	**1,4**
	2	1,7	1,9	2,1	2,6	**1,2**	**1,4**	**1,8**	**2,3**
	3	1,8	2,1	2,6	3,4	**1,3**	**1,7**	**2,2**	**3,1**
Channels 3,7	**1**	1,9	1,9	2,0	2,0	**1,2**	**1,3**	**1,4**	**1,6**
	2	1,9	1,9	2,0	2,2	**1,3**	**1,4**	**1,5**	**1,7**
	3	1,9	2,0	2,2	2,5	**1,3**	**1,5**	**1,8**	**2,2**
Channels 3,5,7	**1**	3,7	3,7	3,7	3,7	**3,4**	**3,4**	**3,4**	**3,4**
	2	3,6	3,5	3,4	3,3	**3,2**	**3,1**	**3,0**	**2,9**
	3	3,6	3,5	3,4	3,3	**3,2**	**3,1**	**3,0**	**3,0**
Channels 3,4,7,8	**1**	2,0	2,0	2,1	2,1	**1,1**	**1,3**	**1,4**	**1,5**
	2	2,0	2,1	2,2	2,3	**1,2**	**1,4**	**1,6**	**1,7**
	3	2,1	2,1	2,2	2,3	**1,3**	**1,5**	**1,6**	**1,8**

Table 13: **Anechoic environment** log-error table with **AIC** $\widehat{N}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with **Φ**.

We see from the above table that $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ is better than $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ in terms of estimation precision. Knowing that $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ is better, we compare the number of signals estimate parameter ($\widehat{N}$) to see which is best in the following table:

Version 2		$\widehat{N}=1$ SNR				$\widehat{N}$ AIC SNR			
	N	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Channels 1,2	**1**	2,5	2,4	2,3	2,3	**1,0**	**1,2**	**1,3**	**1,4**
	2	2,4	2,4	2,4	2,8	**1,2**	**1,4**	**1,8**	**2,3**
	3	2,5	2,5	2,8	3,4	**1,3**	**1,7**	**2,2**	**3,1**
Channels 3,7	**1**	2,4	2,3	2,2	2,2	**1,0**	**1,1**	**1,2**	**1,4**
	2	2,4	2,3	2,2	2,2	**1,0**	**1,1**	**1,3**	**1,5**
	3	2,4	2,3	2,4	2,5	**1,1**	**1,3**	**1,6**	**2,0**
Channels 3,5,7	**1**	2,2	2,1	2,0	1,9	**1,0**	**1,2**	**1,3**	**1,4**
	2	2,1	2,0	2,0	2,1	**1,1**	**1,2**	**1,3**	**1,5**
	3	2,2	2,2	2,3	2,6	**1,1**	**1,3**	**1,5**	**1,7**
Channels 3,4,7,8	**1**	2,0	1,9	1,8	1,8	**1,0**	**1,1**	**1,2**	**1,4**
	2	2,0	1,9	2,0	2,3	**1,1**	**1,2**	**1,4**	**1,6**
	3	2,0	2,1	2,3	2,8	**1,1**	**1,3**	**1,5**	**1,7**

Table 14: **Anechoic environment** log-error table with $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Psi}$ compared with variable and **fixed** $\widehat{\boldsymbol{N}}$.

Clearly the $\widehat{N}$ computed with the AIC is superior to the fixed $\widehat{N}=1$. The only parameter left to vary is the matrix used in the noise whitening process ($\boldsymbol{\Phi}$ or $\boldsymbol{\Psi}$). We will vary theses parameters in the following three tables comparing them with the single channel algorithms:

Channel set	0 dB					5 dB				
	$\boldsymbol{\Phi}$	$\boldsymbol{\Psi}$	MS	Dob.	MMSE	$\boldsymbol{\Phi}$	$\boldsymbol{\Psi}$	MS	Dobl.	MMSE
{1,2}	**1.0**	**1.0**	2.7	3.2	1.8	**1.2**	**1.2**	2.6	3.1	1.9
{3,7}	1.2	**1.0**	2.7	3.2	1.8	1.3	**1.1**	2.6	3.1	1.9
{3,5,7}	3.4	**1.0**	2.7	3.2	1.8	3.4	**1.2**	2.6	3.1	1.9
{3,4,7,8}	1.1	**1.0**	2.7	3.2	1.8	1.3	**1.1**	2.6	3.1	1.9
Channel set	**10 dB**					**15 dB**				
	$\boldsymbol{\Phi}$	$\boldsymbol{\Psi}$	MS	Dobl.	MMSE	$\boldsymbol{\Phi}$	$\boldsymbol{\Psi}$	MS	Dobl.	MMSE
{1,2}	**1.3**	**1.3**	2.6	3.2	2.0	**1.4**	**1.4**	2.6	3.6	2.2
{3,7}	1.4	**1.2**	2.6	3.2	2.0	1.6	**1.4**	2.6	3.6	2.2
{3,5,7}	3.4	**1.3**	2.6	3.2	2.0	3.4	**1.4**	2.6	3.6	2.2
{3,4,7,8}	1.4	**1.2**	2.6	3.2	2.0	1.5	**1.4**	2.6	3.6	2.2

Table 15: **Anechoic environment** log-error table with **AIC** $\widehat{\boldsymbol{N}}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$,and comparing with single channel methods when $\boldsymbol{N}=\mathbf{1}$.

	0 dB					5 dB				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**1.2**	**1.2**	2.7	3.2	1.9	**1.4**	**1.4**	2.6	3.1	2.1
{3,7}	1.3	**1.0**	2.7	3.2	1.9	1.4	**1.1**	2.6	3.1	2.1
{3,5,7}	3.2	**1.1**	2.7	3.2	1.9	3.1	**1.2**	2.6	3.1	2.1
{3,4,7,8}	1.2	**1.1**	2.7	3.2	1.9	1.4	**1.2**	2.6	3.1	2.1
	10 dB					**15 dB**				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**1.8**	**1.8**	2.7	3.3	2.4	**2.3**	**2.3**	3.0	3.9	2.9
{3,7}	1.5	**1.3**	2.7	3.3	2.4	1.7	**1.5**	3.0	3.9	2.9
{3,5,7}	3.0	**1.3**	2.7	3.3	2.4	2.9	**1.5**	3.0	3.9	2.9
{3,4,7,8}	1.6	**1.4**	2.7	3.3	2.4	1.8	**1.7**	3.0	3.9	2.9

Table 16: **Anechoic environment** log-error table with **AIC** $\widehat{N}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$,and comparing with single channel methods when $\boldsymbol{N = 2}$.

Channel set	0 dB					5 dB				
	Φ	**Ψ**	MS	Dobl.	MMSE	**Φ**	**Ψ**	MS	Dobl.	MMSE
{1,2}	**1.3**	**1.3**	2.7	3.2	2.0	**1.7**	**1.7**	2.7	3.1	2.3
{3,7}	1.3	**1.1**	2.7	3.2	2.0	1.5	**1.3**	2.7	3.1	2.3
{3,5,7}	3.2	**1.1**	2.7	3.2	2.0	3.1	**1.3**	2.7	3.1	2.3
{3,4,7,8}	1.3	**1.1**	2.7	3.2	2.0	1.5	**1.3**	2.7	3.1	2.3
Channel set	10 dB					15 dB				
	Φ	**Ψ**	MS	Dobl.	MMSE	**Φ**	**Ψ**	MS	Dobl.	MMSE
{1,2}	**2.2**	**2.2**	2.8	3.4	2.7	**3.1**	**3.1**	3.3	4.1	3.2
{3,7}	1.8	**1.6**	2.8	3.4	2.7	2.2	**2.0**	3.3	4.1	3.2
{3,5,7}	3.0	**1.5**	2.8	3.4	2.7	3.0	**1.7**	3.3	4.1	3.2
{3,4,7,8}	1.6	**1.5**	2.8	3.4	2.7	1.8	**1.7**	3.3	4.1	3.2

Table 17: **Anechoic environment** log-error table with **AIC** $\widehat{N}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$,and comparing with single channel methods when $\boldsymbol{N = 3}$.

We see that our proposed estimate $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$ given by Eq.(4.10) is superior to the MMSE method and other 1-channel methods, except for the 3-channel monaural case (channels 3,5,7). The weaker performance for the 3-channel monaural case might be due to the fact that the coherence matrix has an approximate rank of 1 in that scenario, i.e. the largest eigenvalue of $\boldsymbol{\Phi}$ is much greater than the other two. To our surprise, our proposed coherence based method with $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ is also superior to the MMSE method for channel sets {3,7} and {3,4,7,8} in the anechoic environment, despite the fact that $\boldsymbol{\Phi}$ might be ill-conditioned. In the previous three tables above, there is almost equal performance of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$ or $\boldsymbol{\Psi}$ for the binaural case, but for the other channel sets $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Psi}$ is better than $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ obtained with $\boldsymbol{\Phi}$.

5.3.1.2 Tables of log-error distortion measures in cafeteria environment

The next step is to evaluate the cafeteria environment performances. We will start with the table comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ obtained with $\boldsymbol{\Phi}$ when $\widehat{N} = 1$:

		Version 1 SNR				Version 2 SNR			
	N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
Channels 1,2	**1**	2,8	2,8	2,9	3,2	2,8	2,8	2,9	3,2
	2	2,6	2,7	2,9	3,7	2,6	2,7	2,9	3,7
	3	2,6	2,7	3,1	4,1	2,6	2,7	3,1	4,1
Channels 3,7	**1**	3,5	3,5	3,7	4,3	3,5	3,5	3,7	4,3
	2	3,2	3,2	3,6	4,6	3,2	3,2	3,6	4,6
	3	3,2	3,2	3,7	4,8	3,2	3,2	3,7	4,8
Channels 3,5,7	**1**	**3,5**	**3,6**	**3,8**	**4,5**	4,4	4,2	4,3	4,7
	2	**3,2**	**3,3**	**3,8**	4,9	4,0	3,8	4,0	**4,8**
	3	**3,2**	**3,3**	**3,8**	5,1	4,0	3,8	4,0	**4,9**
Channels 3,4,7,8	**1**	2,6	2,7	3,0	3,7	2,6	2,7	3,0	**3,6**
	2	2,4	2,6	3,2	4,4	2,5	2,6	**3,1**	**4,2**
	3	2,4	2,6	3,3	4,8	2,5	2,6	3,3	**4,6**

Table 18: **Cafeteria environment** log-error table with **fixed** $\widehat{\boldsymbol{N}} = \mathbf{1}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$.

In the above table, we see that for the two channel cases $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ yield the same results as predicted by Eq.(4.11) . The differences occur when the number of channels is greater than 2. The performance of the second estimate is poor compared to the first version for the channel set {3,5,7}.

Let us see what happens when we compute $\widehat{N}$ with the AIC in the following table:

		Version 1 SNR				Version 2 SNR			
	N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
Channels 1,2	**1**	2,2	2,3	2,6	3,0	**1,7**	**1,9**	**2,2**	**2,8**
	2	2,2	2,5	3,0	3,9	**1,8**	**2,1**	**2,7**	**3,7**
	3	2,3	2,6	3,2	4,4	**1,9**	**2,2**	**3,0**	**4,2**
Channels 3,7	**1**	3,1	3,3	3,8	4,5	**2,7**	**3,0**	**3,5**	**4,3**
	2	3,0	3,3	3,8	4,9	**2,7**	**3,0**	**3,7**	**4,8**
	3	3,0	3,3	4,0	5,3	**2,8**	**3,1**	**3,8**	**5,1**
Channels 3,5,7	**1**	6,8	6,4	6,0	5,8	6,8	**6,3**	**5,9**	**5,7**
	2	6,4	5,9	5,4	5,3	6,4	5,9	5,4	**5,2**
	3	6,4	5,8	5,3	5,2	6,4	5,8	5,3	5,2
Channels 3,4,7,8	**1**	3,9	4,0	4,1	4,4	**3,5**	**3,6**	**3,8**	**4,2**
	2	3,8	3,8	4,0	4,5	**3,5**	**3,5**	**3,7**	**4,2**
	3	3,8	3,8	4,1	4,6	**3,5**	**3,6**	**3,8**	**4,5**

Table 19: **Cafeteria environment** log-error table with **AIC** $\widehat{\boldsymbol{N}}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$ and $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$computed with $\boldsymbol{\Phi}$.

We see from the above table that with the use of the AIC the performance of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ is slightly better than $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(1)}$. In the next table, we will compare the effect of $\widehat{N}$ (fixed or computed with AIC) of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$:

Version 2		$\widehat{N} = 1$ SNR				$\widehat{N}$ AIC SNR			
	N	**0dB**	**5dB**	**10dB**	**15dB**	**0dB**	**5dB**	**10dB**	**15dB**
Channels 1,2	**1**	2,8	2,8	2,9	3,2	**1,7**	**1,9**	**2,2**	**2,8**
	2	2,6	2,7	2,9	3,7	**1,8**	**2,1**	**2,7**	3,7
	3	2,6	2,7	3,1	**4,1**	**1,9**	**2,2**	**3,0**	4,2
Channels 3,7	**1**	3,8	3,8	4,0	4,6	**3,1**	**3,4**	**3,8**	4,6
	2	3,5	3,5	3,8	4,8	**3,1**	**3,3**	3,9	5,0
	3	3,5	3,5	3,9	5,0	**3,1**	**3,4**	4,1	5,3
Channels 3,5,7	**1**	**3,7**	**3,7**	**3,9**	**4,6**	4,9	4,7	4,7	5,0
	2	**3,4**	**3,4**	**3,8**	4,9	4,6	4,4	4,4	**4,8**
	3	**3,4**	**3,4**	**3,9**	5,1	4,6	4,4	4,4	5,0
Channels 3,4,7,8	**1**	**2,7**	**2,8**	**3,0**	**3,7**	3,9	3,9	4,1	4,4
	2	**2,5**	**2,7**	**3,2**	**4,3**	3,8	3,8	4,0	4,5
	3	**2,5**	**2,7**	**3,3**	4,7	3,8	3,9	4,1	4,7

Table 20: **Cafeteria environment** log-error table with $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Psi}$ compared with **variable** and **fixed** $\widehat{\boldsymbol{N}}$.

We see that depending on the channel set, the estimator of $\widehat{N}$ affects the precision of the PSD estimate very differently as opposed to the same comparison in the anechoic environment. The following three tables will present the log-errors of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\widehat{N}$ and with either $\boldsymbol{\Phi}$ or $\boldsymbol{\Psi}$. Additionally the single channel methods results are shown:

	0 dB					5 dB				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**1.7**	**1.7**	4.1	4.2	2.5	**1.9**	**1.9**	4.0	4.1	2.6
{3,7}	2.7	3.1	4.1	4.2	**2.5**	3.0	3.4	4.0	4.1	2.6
{3,5,7}	6.8	4.9	4.1	4.2	**2.5**	6.3	4.7	4.0	4.1	2.6
{3,4,7,8}	3.5	3.9	4.1	4.2	**2.5**	3.6	3.9	4.0	4.1	2.6
	10 dB					**15 dB**				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**2.2**	**2.2**	3.8	4.3	2.9	**2.8**	**2.8**	3.9	4.9	3.4
{3,7}	3.5	3.8	3.8	4.3	**2.9**	4.3	4.6	3.9	4.9	**3.4**
{3,5,7}	5.9	4.7	3.8	4.3	**2.9**	5.7	5.0	3.9	4.9	**3.4**
{3,4,7,8}	3.8	4.1	3.8	4.3	**2.9**	4.2	4.7	3.9	4.9	**3.4**

Table 21: **Cafeteria environment** log-error table with **AIC** $\widehat{\boldsymbol{N}}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$,and comparing with single channel methods when $\boldsymbol{N = 1}$.

	0 dB					5 dB				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**1.8**	**1.8**	4.0	4.0	2.6	**2.1**	**2.1**	3.8	3.9	2.9
{3,7}	2.7	3.1	4.0	4.0	**2.6**	3.0	3.3	3.8	3.9	**2.9**
{3,5,7}	6.4	4.6	4.0	4.0	**2.6**	5.9	4.4	3.8	3.9	**2.9**
{3,4,7,8}	3.5	3.8	4.0	4.0	**2.6**	3.5	3.8	3.8	3.9	**2.9**
	10 dB					**15 dB**				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**2.7**	**2.7**	3.8	4.2	3.5	**3.7**	**3.7**	4.2	5.3	4.4
{3,7}	3.7	3.9	3.8	4.2	**3.5**	4.8	5.0	**4.2**	5.3	4.4
{3,5,7}	5.4	4.4	3.8	4.2	**3.5**	5.2	4.8	**4.2**	5.3	4.4
{3,4,7,8}	3.7	4.0	3.8	4.2	**3.5**	4.2	4.5	**4.2**	5.3	4.4

Table 22: **Cafeteria environment** log-error table with **AIC** $\widehat{\boldsymbol{N}}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$,and comparing with single channel methods when $\boldsymbol{N = 2}$.

	0 dB					5 dB				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**1.9**	**1.9**	3.9	4.0	2.6	**2.2**	**2.2**	3.7	3.8	3.0
{3,7}	2.8	3.1	3.9	4.0	**2.6**	3.1	3.4	3.7	3.8	**3.0**
{3,5,7}	6.4	4.6	3.9	4.0	**2.6**	5.8	4.4	3.7	3.8	**3.0**
{3,4,7,8}	3.5	3.8	3.9	4.0	**2.6**	3.6	3.9	3.7	3.8	**3.0**
	10 dB					**15 dB**				
Channel set	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**	**Φ**	**Ψ**	**MS**	**Dobl.**	**MMSE**
{1,2}	**3.0**	**3.0**	3.8	4.2	3.7	**4.2**	**4.2**	4.4	5.3	4.8
{3,7}	3.8	4.1	3.8	4.2	**3.7**	5.1	5.3	**4.4**	5.3	4.8
{3,5,7}	5.3	4.4	3.8	4.2	**3.7**	5.2	5.0	**4.4**	5.3	4.8
{3,4,7,8}	3.8	4.1	3.8	4.2	**3.7**	4.5	4.7	**4.4**	5.3	4.8

Table 23: **Cafeteria environment** log-error table with **AIC** $\widehat{N}$ comparing $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ computed with **Φ**, **Ψ** ,and comparing with single channel methods when $\boldsymbol{N} = \mathbf{3}$.

The important observation from the above three tables is that for the binaural setting, the $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(2)}$ estimate based on either **Φ**, **Ψ** is good, but it fails to be the best method for the other channel sets, in contrast to the anechoic environment results. This might be due to 2 reasons:

- There is a mismatch of the matrix model used in the whitening process, that is, **Φ** or **Ψ** are not entirely reliable. This is due to the fact that the matrices used in the noise whitening process Eq.(4.1) are computed in an anechoic environment and are not suited entirely for the cafeteria environment.
- The log-error measures a distance from the estimated diffuse noise PSD to the true noise PSD. However, the true noise PSD computed has some non-diffuse components (some directional sources are present in the noise recordings being used in the cafetaria environment, which is not entirely diffuse). In the estimation algorithm the non-diffuse noise will be treated as a source and hence will not contribute to the noise PSD matrix estimate. Therefore, the log-error measure should increase.

5.3.2 Selected PSDs and $\widehat{N}$ graphics

In this section we shall show some PSDs in reference to the first channel of a selected set at 2500Hz. For example, if $\mathcal{M} = \{3,7\}$, the reference channel will be 3. All the shown estimated PSDs are computed using Version 2 of the proposed algorithm with fixed or AIC $\widehat{N}$ used. The noise PSD using $\widehat{N} = 1$ is shown in dashed purple with label $\Gamma_{\eta}^{N@1}$. The other dashed green curve is the other noise PSD that depends on the AIC $\widehat{N}$, which is labelled $\Gamma_{\eta}^{N_a}$. In red is the noisy signal PSD and in black the true noise PSD labelled Γ_{η}^{true}. The first scenario shown is the 4-channel binaural setting with three sources at 15dB in the anechoic environment.

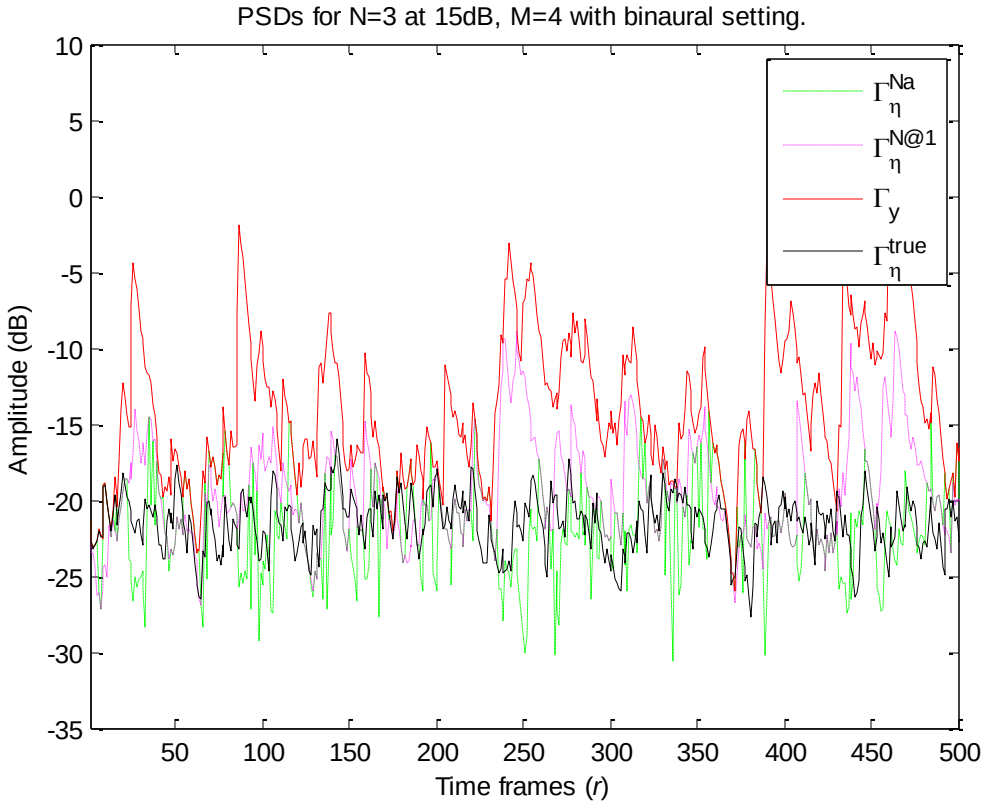

Figure 13: PSD estimates for N=3 at 15dB, M=4 Binaural setting. The noisy signal, the noise for $\widehat{N} = 1$, the noise with the AIC $\widehat{N}$ (Na), and the true noise PSDs are shown respectively in solid red, dashed purple, dashed green, and solid black lines.

We can see for this scenario that the AIC estimation of $\widehat{N}$ can lead to significant improvements, since in this case the actual number of sources is N=3. The following figure illustrates the result of the AIC estimation of $\widehat{N}$ for the same setup.

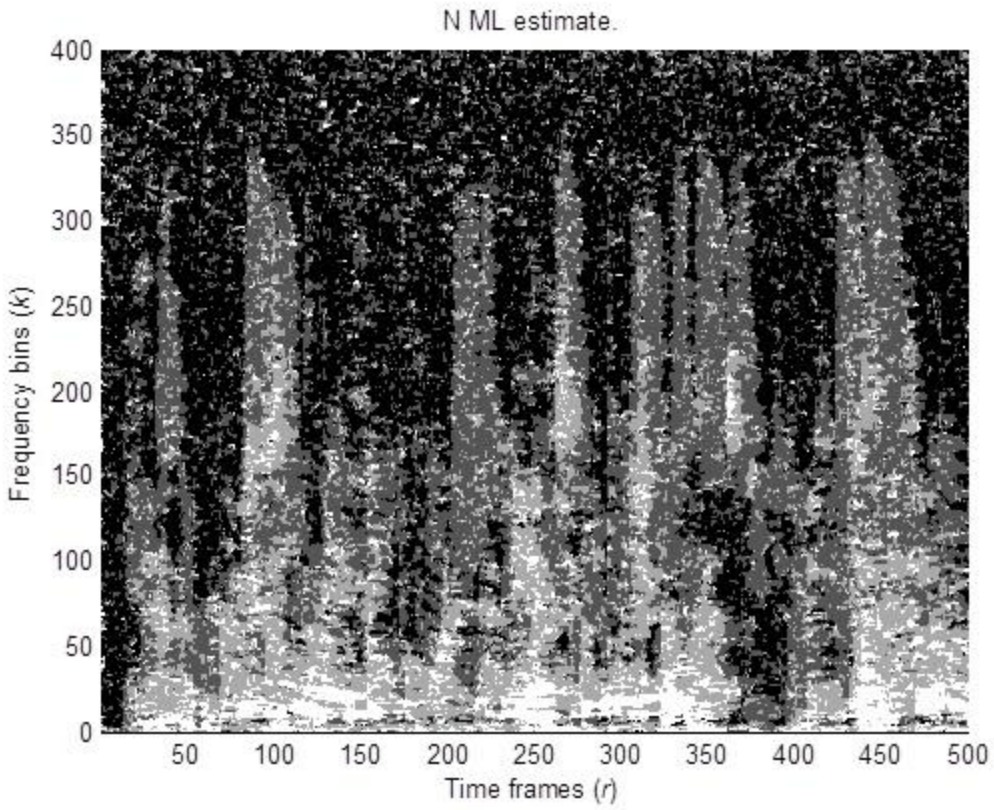

Figure 14: N estimate for N=3 at 15dB, M=4 Binaural setting in **Anechoic environment**.

The above picture shows values of N computed using the AIC criterion. Since $M = 4$ there are four possible values $\widehat{N}$ can take, i.e., 0 up to 3. Black corresponds to $\widehat{N} = 0$ and white corresponds to $\widehat{N} = 3$. Other shades of gray lie between those limit values. There are wave like patterns in the dark regions which obviously correspond to estimation errors (false alarms, too many sources detected). We also clearly see the regions where the sources power is concentrated.

The next setting corresponds to the monaural case, N=1 at 15dB, dual-channel monaural setting in anechoic environment.

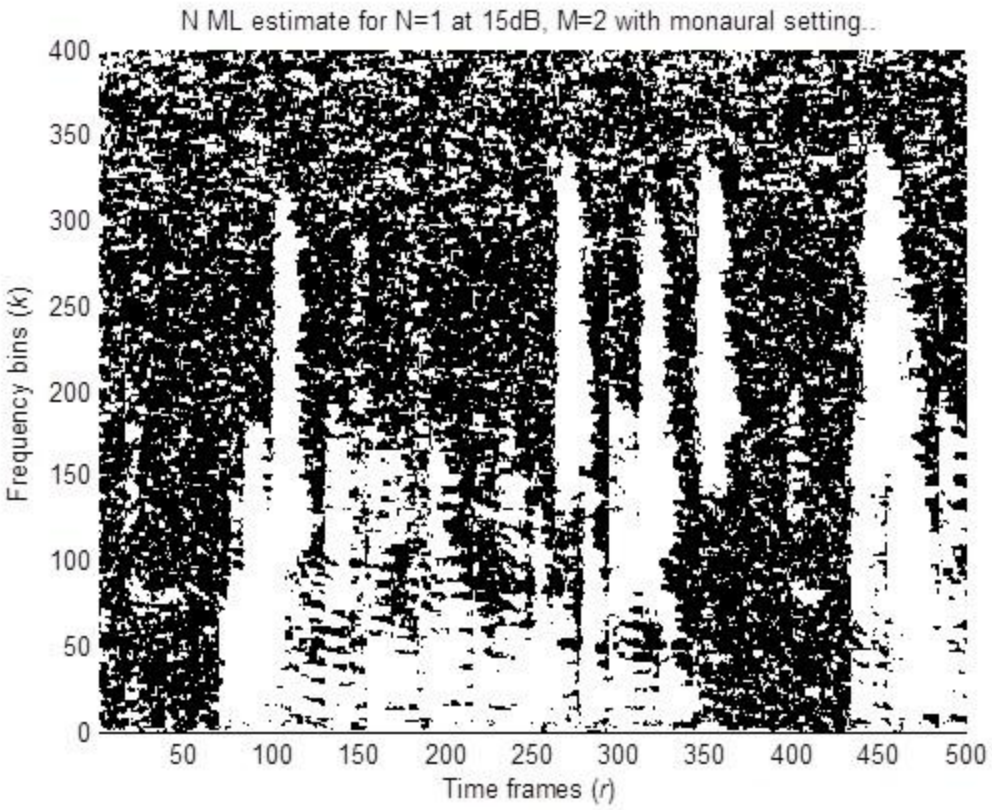

Figure 15: Number of sources estimates ($\widehat{N}$) for N=1 at 15dB, M=2 Monaural setting, Anechoic environment.

Notice how on all AIC estimates of N there seems to be a uniform distribution of clusters of false alarm N estimated values (white spots in black region outside of the speech presence time-frequency region). We also can notice high variations of the noise PSD estimates using AIC $\widehat{N}$. The high variations may be caused by theses rapid change in $\widehat{N}$. Possibly a median filter could be used to remove such outlying PSD values.

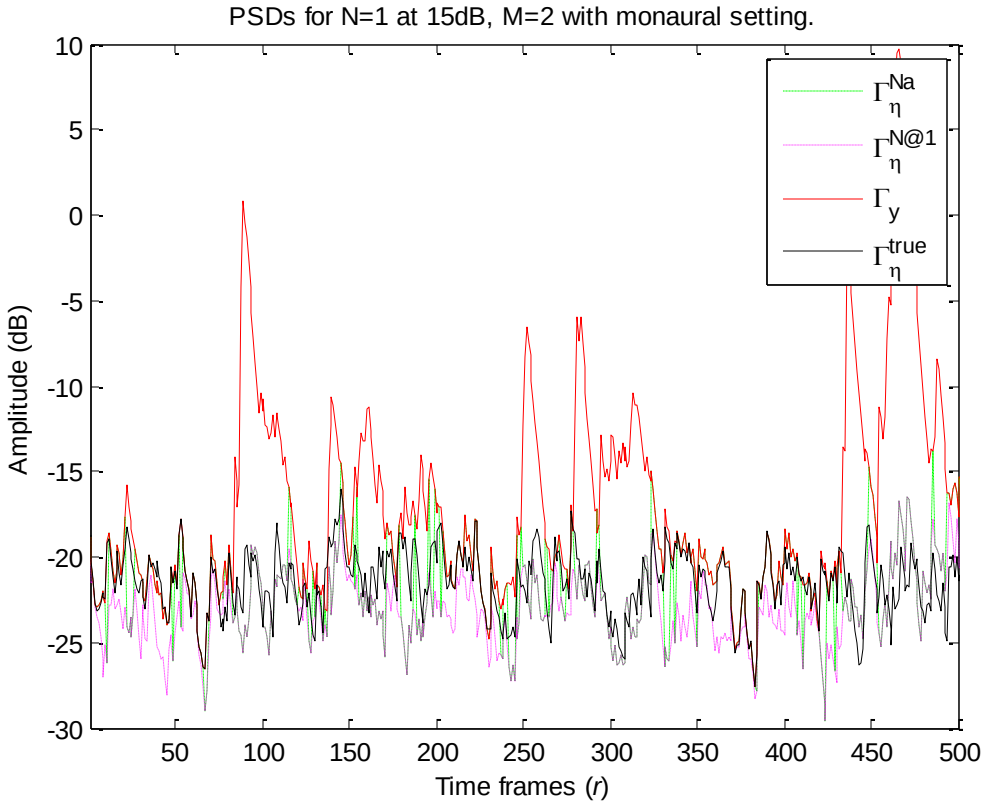

Figure 16: PSD estimates for N=1 at 15dB, M=2 in the Monaural setting (Channel set {3,7}). The PSDs of the noisy signal, the noise for $\widehat{N} = 1$, the noise for the AIC $\widehat{N}$ (Na), and the true noise are shown respectively in solid red, dashed purple, dashed green, and solid black lines.

We observe from Figure 16 that in this case since the actual value of N is 1 the fixed method using $\widehat{N} = 1$ produced roughly the same performance as the method with the AIC estimation of $\widehat{N}$.

5.3.3 Adaptive algorithms potential

We will evaluate the potential of the estimate $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(3)}$ (section 4.1.3 and 4.2.3) by inspecting if the Euclidean distance between $\widehat{\boldsymbol{\Phi}}$ and $\boldsymbol{\Phi}$ tends to diminish and stabilize i.e. converge. The initial condition will be the identity matrix. Note that the Euclidean distance will depend on time and frequency. As mentioned in section 4.1.3and 4.2.3, the algorithm will probably not converge if for the simple fact that by using the ML estimate of N we need a good estimate of $\boldsymbol{\Phi}$, but $\boldsymbol{\Phi}$ itself needs a good estimate of N. Errors in estimation of both parameters will propagate into each other effectively jeopardizing any possible stable convergence. The way around this is by estimating N independently of $\boldsymbol{\Phi}$. But this can be a topic by itself so instead of developing algorithms to estimate N, we will use Eq.(4.15) to model an ideal highly reliable $\widehat{N}$ (that cannot be used in practice). The following figures will show the reliable $\widehat{N}$ and the time-frequency matrix Euclidean norm respectively. We will also compare log-errors measured on two types of $\widehat{\boldsymbol{\Gamma}}_{\eta}^{(3)}$. One will have a reliable $\widehat{N}$ and the other will be using the practical $\widehat{N}$.

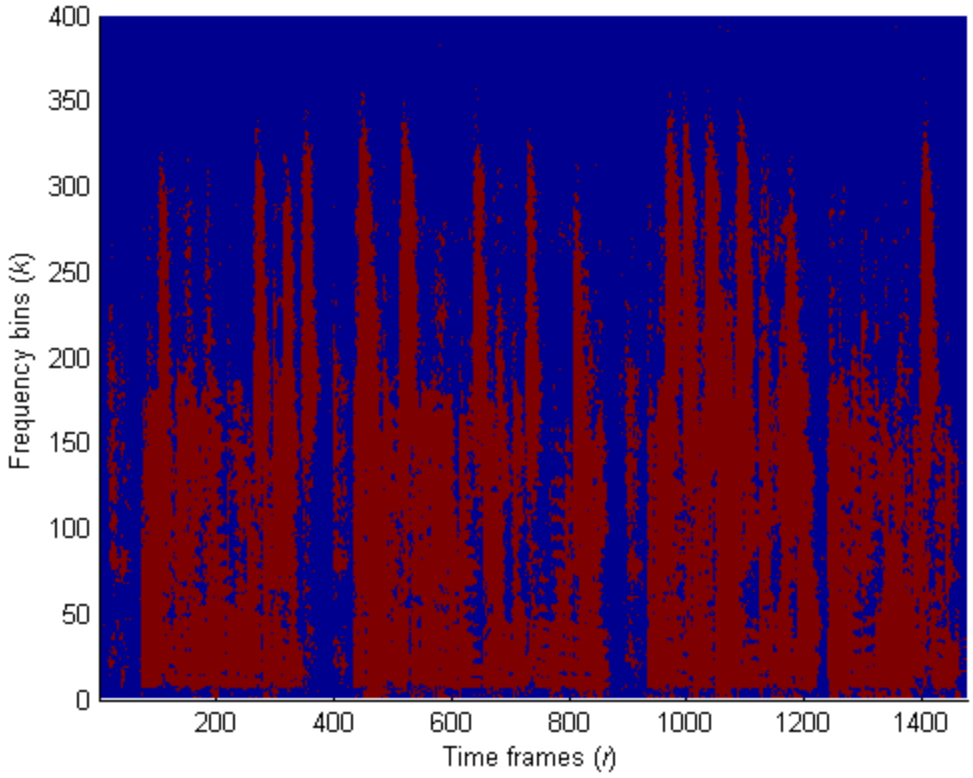

Figure 17: N estimate in **anechoic environment** for the 2-channel binaural setting having one source with 15dB SNR.

We immediately recognise above the pattern of a speech PSD. We see that there is very little false alarm clusters in the regions where *N=0* estimates are located. The red colored region is for *N=1* estimates and the blue region is for *N=0* estimates. The *N=1* estimates indicate the presence of speech, which is similar to a frequency-dependent VAD.

Log-error	
reliable $\widehat{N}$	**unreliable $\widehat{N}$**
1.8	3.6

Table 24: Log-error of third estimate using reliable or unreliable $\widehat{N}$ in **anechoic environment** for the 2-channel binaural setting having source with 15dB of SNR.

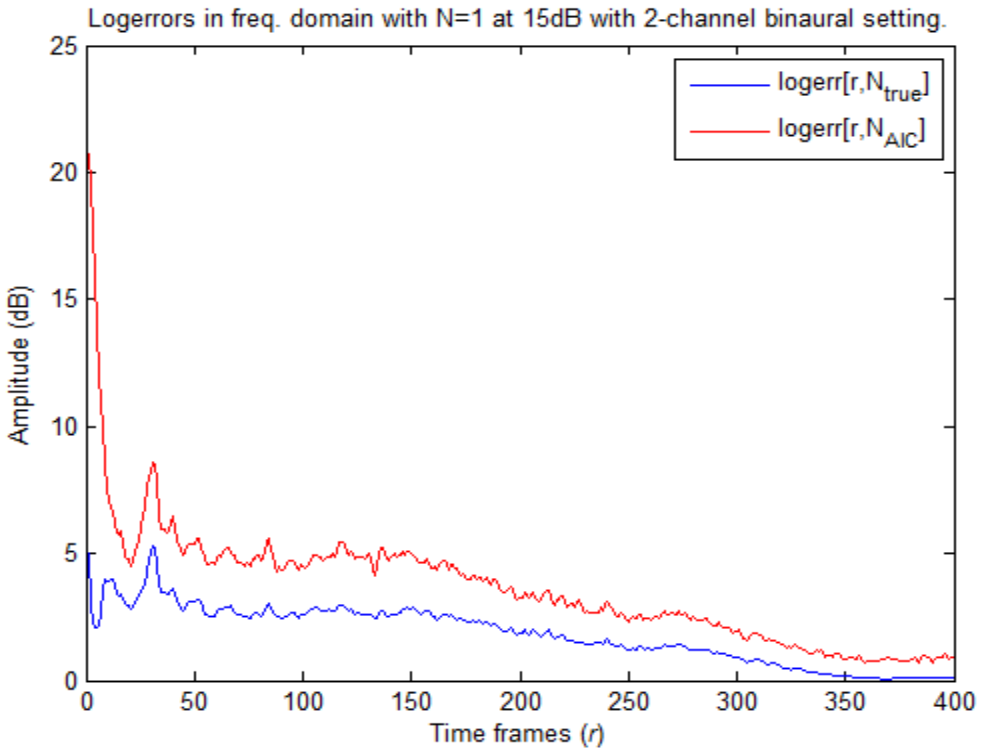

Figure 18: Log-errors in frequency domain of the third estimate with reliable and unreliable $\widehat{N}$ in **anechoic environment** for the 2-channel binaural setting having source with 15dB of SNR.

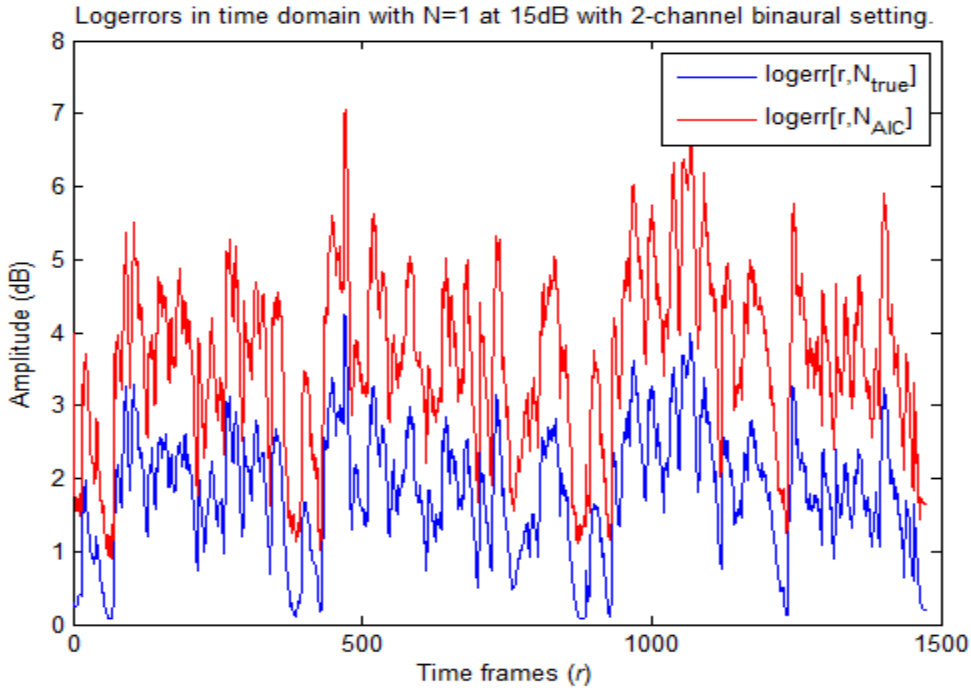

Figure 19: Log-errors in the time domain of the third estimate with reliable and unreliable $\widehat{N}$ in **anechoic environment** for the 2-channel binaural setting having source with 15dB of SNR.

The two graphics above and table 16 show that indeed when we use a reliable $\widehat{N}$ the cost can be greatly reduced. In this case the cost is reduced by a factor of 2. It is likely that this is also true for the other estimator $\mathbf{\Gamma}_{\eta}^{(1)}$ and $\mathbf{\Gamma}_{\eta}^{(2)}$ but to a lesser degree since the coherence matrix does not risk of changing in a static environment. In fact estimating $\widehat{N}_{AIC}$ over a fixed $\widehat{N}$ value is already a reliability improvement and it leads (in general) to a precision improvement shown by the log-error tables 9-15.

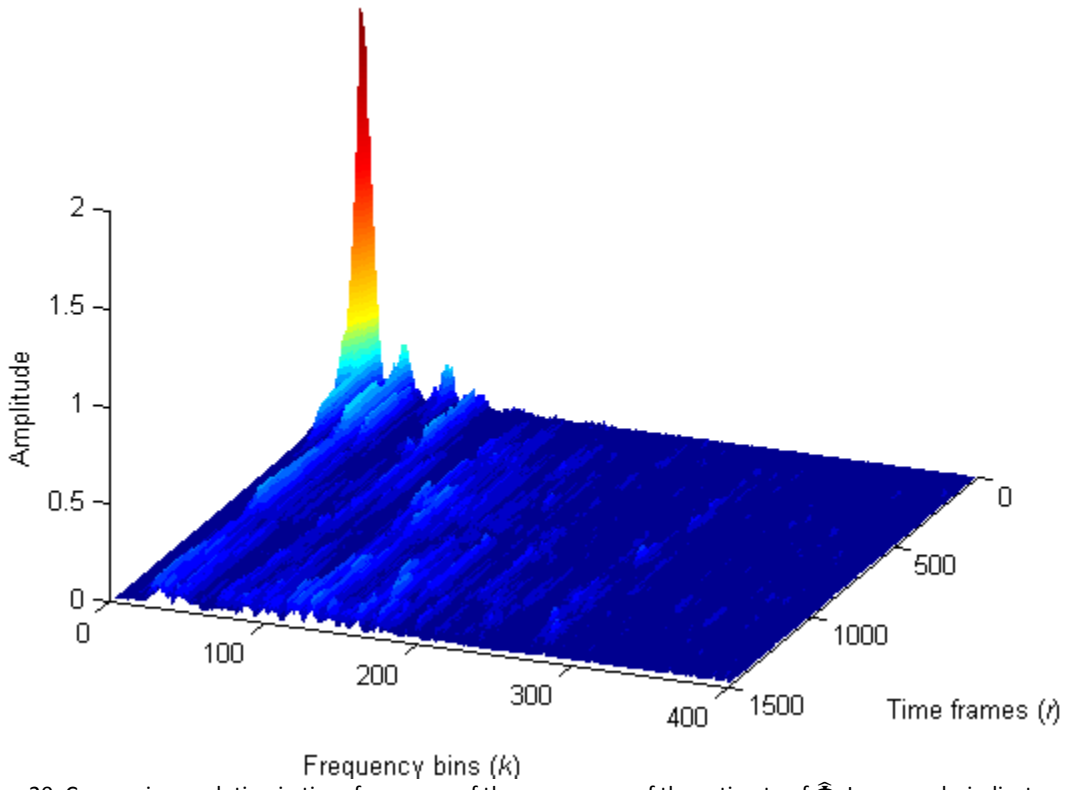

Figure 20: Converging evolution in time-frequency of the error norm of the estimate of $\widehat{\mathbf{\Phi}}$. Large peaks indicate a large error, and we see that the initial peaks have on average decreased.

From the above graphic we see that the initial lobes tend to diminish across time even when there is signal presence. The initial lobes are caused by the difference between the initial condition (initial guess for the coherence) and the true coherence matrix. The initial condition for this case is chosen to be the identity matrix. The above figure tells us that since there is convergence we are able to estimate instantaneous coherences $\mathbf{\Phi}'$ and accumulate these to form an estimate of the environment coherence $\widehat{\mathbf{\Phi}}$. The signal presence $\widehat{N}[r,k] = 1$ just slows down the convergence progression. When $M = 2$ and $\widehat{N}[r,k] = 1$ it even halts the convergence. When there is an error in $\widehat{N}[r,k]$ it might possibly lead to a divergence. We shall next show how the convergence goes when effectively we use the maximum likelihood estimate of $\widehat{N}$.

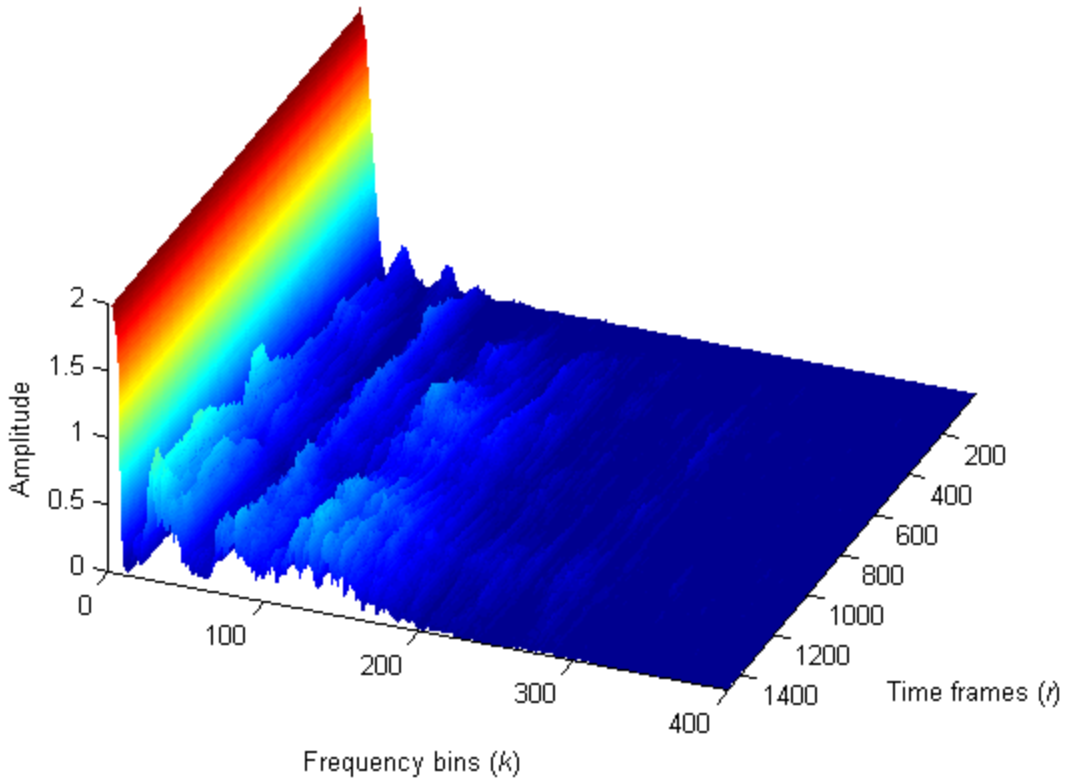

Figure 21: Diverging evolution in time-frequency of the error norm of the estimate of $\widehat{\boldsymbol{\Phi}}$**.** Large peaks indicate a large error, and we see that the initial peaks do not converge to 0.

We can deduce from the above figure that the errors in $\widehat{N}$ will lead to source coherences leaking into the noise coherence estimate, leading to a useless estimator. We know that the source leaked because the energy of the speech is mainly concentrated in low frequencies, i.e., the region where the divergence occurs. This means that from our initial guess $\boldsymbol{I}$ we are unable to obtain good instantaneous coherences estimates and therefore it isn't possible to accumulate them to get a reliable estimate of $\boldsymbol{\Phi}$. Nonetheless there is a high potential of improvement because of the promising converging Euclidean distance. In contrast, the inhomogeneous problem seems much harder to improve primarily because $\boldsymbol{\Psi}$ has no fixed points compared to $\boldsymbol{\Phi}$ which always has unit diagonal elements. The study on the third estimator using $\boldsymbol{\Psi}$ will not be taken any further for this reason.

Chapter 6 Conclusion

A general framework for the multichannel diffuse noise auto-PSD estimation has been successfully put forth in Chapter 2. This framework is applied in Chapter 4 where we developed eigenvalue algorithms and closed form analytical solutions for the inhomogeneous and homogeneous diffuse noise field.

In Chapter 2, in order to further develop the framework, we make an important distinction between coherences and correlations that leads to two similar families of algorithms. Correlation type algorithms are aimed at resolving inhomogeneous noise field PSDs and coherence type algorithms aim at resolving the homogeneous noise fields. It is shown that for either model of noise field, it is possible to estimate the noise PSD of any diffuse noise field.

6.1 Thesis new contributions

The thesis new contributions are listed here:

1. The first new contribution is the analytic model for the coherence function in a spherical isotropic diffuse noise field in the presence of a rigid sphere in section 2.3.1.2, that is Eq.(2.61). The analytical model needs yet to be tested eventually to prove its validity experimentally.
2. The matrix performance measure in Eq.(2.107) in section 2.4.2.4 is a multichannel generalization of the log-error measure often used in the literature, e.g. in [21]. We use it for the performance evaluation of the algorithms in Chapter 5.
3. The algorithm derived in section 4.1.1 already exists under various forms [3, 42], but it is the first time that it is applied in the general framework for acoustic sources in the MIMO perspective under a homogeneous diffuse noise field.
4. The coherence based algorithms in section 4.1.2 and 4.1.3 can be considered completely new. They solve two problems:
   - Modeling the noise PSD matrix with different noise auto-PSDs when the homogeneous noise field model used inherently supposes that the auto PSDs are equal (noise field model mismatch).
   - If there is no noise field model mismatch, the auto-PSDs are still almost surely not equal in practice (although the differences can be small), and the coherence matrix used in the algorithm in section 4.1.1 forces the auto-PSDs to be equal.

   The coherence based algorithms in section 4.1.2 and 4.1.3 also use the number of signals estimate to increase the precision of the noise PSD matrix estimate.
5. The algorithm in section 4.1.3 estimates the coherence matrix adaptively. This is needed because noise field coherences or correlations matrices are only valid in a static environment. If change occurs, our premeasured estimates of the coherences or correlations might be rendered obsolete. The algorithm addresses this issue. It was shown in section 5.3.3 that the algorithm of section 4.1.3 needs a proper number of sources estimate in order to work effectively.
6. The algorithm described in 4.2.1 was previously used, albeit not in our context, which can be considered a new contribution.
7. The coherence based algorithm in section 4.2.2 is also new. The algorithm solves the problem of increasing the precision of the noise PSD matrix estimate by using noise subspace knowledge.

8. In section 4.2.3 an analog solution to the algorithm provided in 4.1.3 is shown, but the correlation matrix adaptive estimation is not stable. Further study is needed in order to find a stabilising solution in order to make the algorithm in section 4.2.3 work correctly.
9. It is shown in section 4.3.1 that when the noise auto PSDs are different (e.g., in a inhomogeneous noise field) and we have the coherence matrix at our disposal, it is possible to find the single source PSD algebraically under the conditions that the transfer functions of the source are known with a binaural scenario. The solution is unusable in practice since it is too sensitive to errors and it was not designed for sample PSD matrices. There is a door open for possible improvements (for example using regularisation) to make the solution usable although in a modified form.
10. Chapter 5 confirms the applicability of the new techniques. We have found that it is even possible to take advantage of very closely spaced microphones, in particular when the correlation type algorithms are used. In fact, it is shown in Chapter 5 that the correlation based algorithms produce the best cost function scores in several setups compared to popular single-channel algorithms discussed in Chapter 3, in particular under anechoic environments. Other coherence based developed algorithms clearly outperform single-channel algorithms when we use a binaural setting in the highly non-stationary noise environment of the cafeteria. Multiple parameters such as the number of sources and the SNR were varied to pinpoint which algorithm works better in which scenario. This vast variety of possible combinations permitted us to learn even about pre-existing single channel methods.
11. The algorithms of sections 4.1.2 and 4.2.2 were presented in an article "Short-time multichannel noise correlation matrix estimators for acoustic signals" to be published in the 2014 HSCMA conference (Nancy, France, May 2014).

6.2 Future work

Here is a list of possible future work:

1. It was shown in Chapter 5 that the section 4.1.3 algorithm only works if we have a reliable estimate of the number of sources. Additionally, the simulations show that if the estimate of the number of sources is good, then the precision of the noise PSD estimate (algorithm in sections 4.1.1,4.1.2,4.1.3,4.2.1,and 4.2.2) increases. These facts are definitely a motivation to research for source number estimation schemes as future work to permit a noise PSD estimation algorithm robust in all environments. This is a model order estimation problem and the models need to work with the following specifications:
   - The noisy signal PSD matrix is computed recursively and not with the moving average technique.
   - Since we are interested in short-time estimation, the estimator needs to be robust with "short data records" or "low-support".
   - Possibly a Monte Carlo method algorithm might offer a robust criterion based on an appropriate statistical (or empirical) model for the joint distribution of the eigenvalues.
2. As mentioned in the contributions, the adaptive correlation estimation technique (section 4.2.3) needs to be reworked so that it becomes more stable.
3. Develop new multichannel speech enhancement algorithms by generalizing existing single-channel techniques (e.g. spectral subtraction) into the multichannel domain. For example, the MMSE algorithm noise PSD expression can be easily translated into multiple channels using Bayesian inference, but the difficult part would be to model the SNR definitions in the multichannel case, for example the decision directed SNR estimation, the maximum likelihood SNR, etc.

4. Investigate the psychoacoustic qualities of the different multichannel performance measures when the estimate is an “optimal” one with respect to the performance measure. This includes doing a multichannel noise reduction of the measured signals to hear, for example if the noise PSD matrix preserves the spatial information of the sources. Also this will need to be compared with other single channel algorithms to see how it performs. Noise reduction holds a variety of techniques and many have parameters that need to be tuned for a specified performance. However, many of the classical noise reduction techniques are only defined in the single channel domain.

Appendices

A. On the periodogram distribution

We are interested in finding the distribution of the magnitude squared of a discrete STFT of a given signal ($|Y[r,k]|^2$), given that it's real and imaginary parts are two zero mean real independent Gaussian distributed random variables of variance $\frac{\sigma^2}{2}$. The periodogram real and imaginary parts are denoted by Y_R and Y_I respectively. We are interested of finding $f_{|Y[r,k]|^2}(z)$ such that $|Y[r,k]|^2 = Y_R^2 + Y_I^2$ where the variables are jointly Gaussian:

$$\begin{pmatrix} Y_R \\ Y_I \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\sigma^2}{2} I_{2\times 2} \right).$$

Noting that Y_R^2 and Y_I^2 are independent and identically distributed (i.i.d), we can compute the pdf of $|Y[r,k]|^2$ from the characteristic function of one of the variables.

$$f_{Y_R^2}(t) = \frac{f_X(\sqrt{t}) + f_X(-\sqrt{t})}{2\sqrt{t}} u(t) = \frac{1}{\sqrt{\pi\sigma^2 t}} e^{-\frac{t}{\sigma^2}} u(t),$$

$$\Phi_{Y_R^2}(j\omega) = \mathcal{E}\left(e^{j\omega t}\right) = \int_0^\infty \frac{1}{\sqrt{\pi\sigma^2 t}} e^{-\left(\frac{1}{\sigma^2} - \jmath\omega\right)t}\, dt = \frac{\sqrt{\pi}}{\sqrt{\pi\sigma^2\left(\frac{1}{\sigma^2} - \jmath\omega\right)}},$$

$$\Phi_{|Y[r,k]|^2}(j\omega) = \Phi_{Y_R^2}(j\omega)\Phi_{Y_I^2}(j\omega) = \frac{1}{\sigma^2\left(\frac{1}{\sigma^2} - \jmath\omega\right)},$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-\jmath\omega t}}{\sigma^2\left(\frac{1}{\sigma^2} - \jmath\omega\right)} d\omega = \frac{1}{\sigma^2} e^{-\frac{t}{\sigma^2}} u(t) = f_{|Y[r,k]|^2}(t).$$

If we now consider the average of L independent exponentially distributed variables $Z_i = \frac{1}{L}|Y[r+i,k]|^2$ for $i = [1, L]$:

$$\frac{1}{L}\sum_{i=0}^{L-1} |Y[r+i,k]|^2 = \sum_{i=0}^{L-1} Z_i = P.$$

$$f_{Z_i}(t) = \frac{L}{\sigma^2} e^{-\frac{Lt}{\sigma^2}} u(t),$$

where **A.1**

$$u(t) = \begin{cases} 0; t \le 0 \\ 1; t > 0 \end{cases}.$$

Since the variables Z_i are i.i.d.,

$$\Phi_P(j\omega) = \left(\Phi_{Z_i}(j\omega)\right)^L = \left(\frac{\sigma^2}{L}\left(\frac{L}{\sigma^2} - j\omega\right)\right)^{-L},$$

$$\Phi_P(j\omega) = \left(1 - j\omega\frac{\sigma^2}{L}\right)^{-L}. \quad \textbf{A.2}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-\jmath\omega t}}{\left(\frac{\sigma^2}{L}\left(\frac{L}{\sigma^2} - \jmath\omega\right)\right)^L} d\omega = f_P(t),$$

$$f_P(t) = \frac{t^{L-1}}{\frac{\sigma^{2L}}{L^L}\Gamma(L)} e^{-\frac{Lt}{\sigma^2}} u(t). \quad \textbf{A.3}$$

The expected value of the random variable P yields the variance σ^2. The mean square value is given by:

$$\mathcal{E}(P^2) = \frac{L^2 + L}{\left(\frac{L}{\sigma^2}\right)^2} = \sigma^4 + \frac{\sigma^4}{L}.$$

Thus the variance is

$$var(P) = \mathcal{E}(P^2) - \mathcal{E}(P)^2 = \frac{\sigma^4}{L}.$$

The cumulative distribution function of the variable P is given by:

$$\int_0^t \frac{u^{L-1}}{\frac{\sigma^{2L}}{L^L}\Gamma(L)} e^{-\frac{Lu}{\sigma^2}} du = \int_0^{\frac{\sigma^2}{L}t} \frac{u^{L-1}}{\Gamma(L)} e^{-u} du = \frac{\gamma\left(L, \frac{\sigma^2}{L}t\right)}{\Gamma(L)} = F_P(t).$$

Where $\gamma(n,x)$ is the lower incomplete gamma function[16]. If n is an integer, $\gamma(n,x)$ can be expressed by the following formulas:

$$\gamma(n,x) = \int_0^x u^{n-1} e^{-u} du,$$

$$\frac{\gamma(n,x)}{\Gamma(n)} = 1 - e^{-x}\sum_{m=0}^{n-1}\frac{x^m}{m!} = 1 - \frac{\Gamma(n,x)}{\Gamma(n)},$$

where,

$$\Gamma(n,x) = \int_x^{\infty} u^{n-1} e^{-u} du.$$

B. On the moving average periodogram estimate minimum distribution.

Suppose there are D independent random variables $\{P_1, P_2, \dots, P_D\}$ that are identically distributed to P. Then, the minimum of the set, that is $P_{min} = \min(P_1, P_2, \dots, P_D)$ has the following pdf [8]

$$f_{P_{min}}(t) = D\big(1 - F_P(t)\big)^{D-1} f_P(t), \tag{B.1}$$

where $f_P(t)$ is the probability density function defined in Eq.(A.1).

P_{min} has the following cdf

$$F_{P_{min}}(t) = \int_0^t D\big(1 - F_P(u)\big)^{D-1} f_P(u)du = 1 - \big(1 - F_P(t)\big)^D. \tag{B.2}$$

The expected value of P_{min} is

$$\mathcal{E}(P_{min}) = \int_0^\infty tD\big(1 - F_P(t)\big)^{D-1} f_P(t)dt = -t\big(1 - F_P(t)\big)^D\Big|_0^\infty + \int_0^\infty \big(1 - F_P(t)\big)^D dt,$$

$$\mathcal{E}(P_{min}) = \int_0^\infty \big(1 - F_P(t)\big)^D dt. \tag{B.3}$$

$$\mathcal{E}(P_{min}) = \int_0^\infty \left(1 - \frac{\gamma\left(L, \frac{\sigma^2}{L}t\right)}{\Gamma(L)}\right)^D dt = \int_0^\infty e^{-\frac{\sigma^2}{L}Dt}\left(\sum_{m=0}^{L-1}\frac{\sigma^{2m}t^m}{m!\,L^m}\right)^D dt.$$

For the special case $L = 1$;

$$\mathcal{E}(P_{min}) = \int_0^\infty e^{-\sigma^2 Dt}dt = \frac{1}{\sigma^2 D}.$$

For the special case $L = 2$;

$$\mathcal{E}(P_{min}) = \int_0^\infty e^{-\frac{\sigma^2}{2}Dt}\left(1 + \frac{\sigma^2}{2}t\right)^D dt = \frac{2}{\sigma^2}\int_0^\infty e^{-Dt}(1+t)^D dt = \frac{2}{\sigma^2}e^D\int_1^\infty e^{-Dt}t^D dt$$

$$= \frac{2}{\sigma^2 D}D^{-D}e^D\int_D^\infty e^{-t}t^D dt = \frac{2}{\sigma^2 D}\frac{e^D\Gamma(D+1, D)}{D^D}.$$

For the special case $L = 3$;

$$\mathcal{E}(P_{min}) = \frac{3}{\sigma^2}\int_0^\infty e^{-Dt}\sum_{k=0}^{D}\binom{D}{k}t^k\left(1 + \frac{t}{2}\right)^k dt$$

$$= \frac{3}{\sigma^2}\int_0^\infty e^{-Dt}\sum_{k=0}^{D}\binom{D}{k}t^k\left(1 + \frac{t}{2}\right)^k dt = \frac{3}{\sigma^2}\int_0^\infty e^{-Dt}\sum_{k=0}^{D}\binom{D}{k}t^k\sum_{m=0}^{k}\binom{k}{m}\frac{t^m}{2^m}dt$$

$$= \frac{3}{\sigma^2 D} D! \sum_{k=0}^{D} \sum_{m=0}^{k} \frac{(m+k)!}{2^m D^{m+k} (D-k)!\, m!\, (D-m)!}$$

In general,

$$\mathcal{E}(P_{min}) = \frac{L}{\sigma^2 D} g_L(D),$$

where $g_L(D)$ is a rational function of D provided that D and L are positive integers. The function $g_L(D)$ can be numerically calculated via the integral

$$g_L(D) = D \int_0^{\infty} e^{-Dt} \left(\sum_{n=0}^{D} \frac{t^n}{n!} \right)^D dt.$$

C. On the pdf of periodograms done with dependent samples .

Case 1: Periodogram estimate using a moving average and circularly symmetric complex Gaussian variables $\mathrm{Y}[\mathrm{r}, \mathrm{k}]$

We are interested in finding the characteristic function of a periodogram using a moving average.

$$P = \frac{1}{L} \sum_{i=0}^{L-1} |Y[r-i,k]|^2.$$

The i^{th} periodogram real and imaginary part are denoted by $Y_R[r-i,k]$ and $Y_I[r-i,k]$ respectively. We still suppose that imaginary parts are independent from real parts, but between the time samples of $Y[r,k]$ there is a correlation. All the variables are jointly gaussian. If we concatenate the real parts of the samples $Y[r,k]$ over the imaginary parts and normalise the resulting vector by L we get the vector of random variables $\mathcal{X}$:

$$\mathcal{Y}_R = \begin{bmatrix} Y_R[r,k] \\ Y_R[r-1,k] \\ \vdots \\ Y_R[r-L+1,k] \end{bmatrix}, \mathcal{Y}_I = \begin{bmatrix} Y_I[r,k] \\ Y_I[r-1,k] \\ \vdots \\ Y_I[r-L+1,k] \end{bmatrix},$$

$$\mathcal{X} = \frac{1}{\sqrt{L}} \begin{bmatrix} \mathcal{Y}_R \\ \mathcal{Y}_I \end{bmatrix},$$

$$\mathcal{X} \sim N\left(\mathbf{0}, \frac{1}{2L} \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix} \right).$$

The periodogram is thus the quadratic form

$$P = \mathcal{X}^T \mathcal{X} = \frac{1}{L} \sum_{i=0}^{L-1} |Y[r-i,k]|^2.$$

The characteristic function of the random variable P is derived in the following way:

$$f_{\mathcal{X}}(\boldsymbol{x}) = \frac{e^{-L\boldsymbol{x}^T \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{bmatrix} \boldsymbol{x}}}{(2\pi)^L \left| \frac{1}{2L} \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix} \right|^{\frac{1}{2}}} = \frac{L^L}{\pi^L |\Sigma|} e^{-L\boldsymbol{x}^T \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{bmatrix} \boldsymbol{x}},$$

$$\mathcal{E}\left(e^{j\omega \boldsymbol{x}^T\boldsymbol{x}}\right) = \int_{\mathbb{R}^{2L}} e^{j\omega \boldsymbol{x}^T\boldsymbol{x}} \frac{L^L}{\pi^L|\Sigma|} e^{-L\boldsymbol{x}^T\begin{bmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{bmatrix}\boldsymbol{x}} d\boldsymbol{x},$$

$$= \frac{L^L}{\pi^L|\Sigma|} \int_{\mathbb{R}^{2L}} e^{-\boldsymbol{x}^T\left(L\begin{bmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{bmatrix} - j\omega I_{2L\times 2L}\right)\boldsymbol{x}} d\boldsymbol{x}$$

$$= \frac{L^L}{\pi^L|\Sigma|} \int_{\mathbb{R}^{2L}} e^{-\boldsymbol{w}^T\boldsymbol{w}} \left|\begin{bmatrix}L\Sigma^{-1} - j\omega I & 0\\ 0 & L\Sigma^{-1} - j\omega I\end{bmatrix}\right|^{-1/2} d\boldsymbol{w}\,; Im(\omega) > -L\lambda_i^{-1}$$

$$= \frac{L^L}{|\Sigma||L\Sigma^{-1} - j\omega I|} = \frac{1}{\left|\mathrm{I} - j\omega\frac{\Sigma}{L}\right|} = \prod_{i=1}^{L} \frac{1}{\left(1 - j\omega\frac{\lambda_i}{L}\right)}.$$

$$\Phi_P(j\omega) = \prod_{i=1}^{L} \frac{1}{\left(1 - j\omega\frac{\lambda_i}{L}\right)}, \qquad \textbf{C.1}$$

with λ_i equal to the eigenvalues of the covariance matrix Σ. Note that if $\Sigma = \sigma^2\mathrm{I}$ then we get

$$\Phi_P(j\omega) = \left(1 - j\omega\frac{\sigma^2}{L}\right)^{-L}.$$

which is the characteristic function found earlier in Eq.(A.2).

Case 2: Periodogram estimate using a recursion and circularly symmetric complex Gaussian variables Y[r, k].

We are interested in finding the characteristic function of the periodogram estimate computed with a recursion:

$$P = (1-\alpha)\sum_{i=0}^{L-1} \alpha^i |Y[r-i,k]|^2.$$

The same gaussianity assumptions are kept for $\mathcal{Y}_R$ and $\mathcal{Y}_I$. Let $\mathcal{X}$ be the concatenation of $\mathcal{Y}_R$ and $\mathcal{Y}_I$:

$$\mathcal{X} = \begin{bmatrix}\mathcal{Y}_R\\ \mathcal{Y}_I\end{bmatrix}, \mathcal{X} \sim N\left(\mathbf{0}, \frac{1}{2}\begin{bmatrix}\Sigma & 0\\ 0 & \Sigma\end{bmatrix}\right)$$

The periodogram is thus the quadratic form

$$P = \mathcal{X}^T\begin{bmatrix}\mathrm{A} & 0\\ 0 & \mathrm{A}\end{bmatrix}\mathcal{X} = (1-\alpha)\sum_{i=0}^{L-1} \alpha^i |Y[r-i,k]|^2,$$

$$\mathrm{A} = (1-\alpha)\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \alpha & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \alpha^{L-1}\end{bmatrix}.$$

The characteristic function of the random variable *P* is derived in the following way:

$$\mathcal{E}\left(e^{j\omega \boldsymbol{x}^T \boldsymbol{x}}\right) = \frac{1}{\pi^L |\Sigma|} \int_{\mathbb{R}^{2L}} e^{\jmath\omega \boldsymbol{x}^T \begin{bmatrix} \mathrm{A} & 0 \\ 0 & \mathrm{A} \end{bmatrix} \boldsymbol{x}} e^{-\boldsymbol{x}^T \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{bmatrix} \boldsymbol{x}} d\boldsymbol{x},$$

$$= \frac{1}{\pi^L |\Sigma|} \int_{\mathbb{R}^{2L}} e^{-\boldsymbol{x}^T \left(\begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{bmatrix} - \jmath\omega \begin{bmatrix} \mathrm{A} & 0 \\ 0 & \mathrm{A} \end{bmatrix} \right) \boldsymbol{x}} d\boldsymbol{x}$$

$$= \frac{1}{\pi^L |\Sigma|} \begin{vmatrix} \Sigma^{-1} - \jmath\omega \mathrm{A} & 0 \\ 0 & \Sigma^{-1} - \jmath\omega \mathrm{A} \end{vmatrix}^{-1/2} \int_{\mathbb{R}^{2L}} e^{-\boldsymbol{w}^T \boldsymbol{w}} d\boldsymbol{w}\,;\, Im(\omega) > -\lambda_i^{-1}$$

$$\mathcal{E}\left(e^{j\omega \boldsymbol{x}^T \boldsymbol{x}}\right) = \frac{1}{|\Sigma||\Sigma^{-1} - \jmath\omega \mathrm{A}|} = \frac{1}{|\mathrm{I} - \jmath\omega \Sigma \mathrm{A}|} = \prod_{i=1}^{L} \frac{1}{(1 - \jmath\omega\lambda_i)}.$$

$$\Phi_P(j\omega) = \prod_{i=1}^{L} \frac{1}{(1 - \jmath\omega\lambda_i)}, \qquad \text{C.2}$$

where λ_i are the eigenvalues of ΣA. Note that the characteristic function is essentially of the same form as in Eq. (C.1.1)

D. The minimal root of a polynomial of degree 3 with positive zeros.

Let the cubic function $f(\lambda)$ having all real coefficients and positive roots be defined by:

$$f(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d.$$

The polynomial is set to zero, and we divide it by a which is non-zero,

$$\lambda^3 + \frac{b}{a}\lambda^2 + \frac{c}{a}\lambda + \frac{d}{a} = 0.$$

Attempting to "complete the cube" with the substitution $y = \lambda - \frac{b}{3a}$ will yield the depressed cubic:

$$y^3 + py + q = 0.$$

with $= \frac{1}{3a^2}(3ac - b^2)$, and $q = \frac{1}{3^3 a^3}(2b^3 - 9abc + 27a^2 d)$. When all roots are positive, François Viète's[43] line of thought can be followed which consists in coinciding the depressed cubic with the trigonometric identity :

$$\cos^3(\theta) - \frac{3}{4}\cos(\theta) - \frac{1}{4}\cos(3\theta) = 0.$$

Substituting $y = t\cos(\theta)$ with t being a matching parameter to be determined,

$$t^3 \cos^3(\theta) + pt\cos(\theta) + q = 0.$$

Dividing by t^3 we get

$$\cos^3(\theta) + \frac{p}{t^2}\cos(\theta) + \frac{q}{t^3} = 0.$$

Solving for the parameter t we have

$$\frac{p}{t^2} = -\frac{3}{4} \rightarrow t = 2\sqrt{\frac{-p}{3}},$$

and

$$\cos^3(\theta) - \frac{3}{4}\cos(\theta) + \frac{q}{t^3} = 0.$$

Then

$$\frac{q}{t^3} = -\frac{1}{4}\cos(3\theta) \rightarrow \cos(3\theta) = \frac{3q}{2p}\sqrt{\frac{3}{-p}}.$$

Solving for the angle θ we get

$$\theta_k = \frac{1}{3}\operatorname{acos}\left(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\right) + \frac{2\pi k}{3} = \psi + \frac{2\pi k}{3},$$

and back substituting the values to find the roots λ

$$\lambda_k = -\frac{b}{3a} + 2\sqrt{\frac{-p}{3}}\cos\left(\psi + \frac{2\pi k}{3}\right).$$

Looking only at the term where *k=1*, λ_1 is a minimum whenever $\psi \in \left[0, \frac{2\pi}{3}\right]$. It can be derived that for the condition on ψ to work, the discriminant of the cubic $\Delta > 0$. But when $\Delta > 0$ the roots are all real and distinct which is always true by hypothesis. Hence the minimum root of the cubic is given by the formula:

$$\lambda_{min} = -\frac{b}{3a} + 2\sqrt{\frac{-p}{3}}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\right) + \frac{2\pi}{3}\right). \quad \textbf{D.1}$$

Using similar arguments it can be shown that the middle and largest eigenvalue are given by:

$$\lambda_{middle} = -\frac{b}{3a} + 2\sqrt{\frac{-p}{3}}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\right) + \frac{4\pi}{3}\right), \quad \textbf{D.2}$$

$$\lambda_{max} = -\frac{b}{3a} + 2\sqrt{\frac{-p}{3}}\cos\left(\frac{1}{3}\operatorname{acos}\left(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\right)\right). \quad \textbf{D.3}$$

E. Coefficient formula.

Let **B** and $\boldsymbol{A}$ be positive definite $M \times M$ Hermitian matrices. The polynomial $\det(\lambda\boldsymbol{B} - \boldsymbol{A})$ can be expanded as:

$$\det(\lambda\boldsymbol{B} - \boldsymbol{A}) = \lambda^M c_M + \lambda^{M-1} c_{M-1} + \cdots + c_0$$

$$c_0 = (-1)^M \det(\boldsymbol{A}), and\ c_M = \det(\boldsymbol{B})$$

$$c_{M-m} = (-1)^m \sum_{1 \le i_1 < \cdots < i_m \le M} \det\left[\underset{i_1,\ldots,i_m}{\mathrm{col}}\, \boldsymbol{A} \to \boldsymbol{B}\right]$$

where $\mathrm{col}_{i_1,\ldots,i_m}\, \boldsymbol{A} \to \boldsymbol{B}$ is a matrix formed by inserting the $i_1^{th}, \ldots, i_m^{th}$ columns of $\boldsymbol{A}$ into $\boldsymbol{B}$. If one needs to use a vectorised formula to compute the coefficients, the Leibnitz formula is recommended to be used:

$$\det(\boldsymbol{A}) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

References

[1] M. Jeub, C. M. Nelke, H. Krüger, C. Beaugeant, P. Vary and C. Herglotz, "Robust dual-channel noise power spectral density estimation," *Proceedings of European Signal Processing Conference (EUSIPCO),* pp. 2304-2308, 2011.

[2] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant and P. Vary, "Noise reduction for dual-microphone mobile phones exploiting power level differences," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference On,* 2012, pp. 1693-1696.

[3] A. H. Kamkar-Parsi and M. Bouchard, "Improved Noise Power Spectrum Density Estimation for Binaural Hearing Aids Operating in a Diffuse Noise Field Environment," *Audio, Speech, and Language Processing, IEEE Transactions On,* vol. 17, pp. 521-533, 2009.

[4] R. C. Hendriks and T. Gerkmann, "Noise Correlation Matrix Estimation for Multi-Microphone Speech Enhancement," *Audio, Speech, and Language Processing, IEEE Transactions On,* vol. 20, pp. 223-233, 2012.

[5] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous Binaural Target PSD Estimation for Hearing Aid Noise Reduction in Complex Acoustic Environments," *Instrumentation and Measurement, IEEE Transactions On,* vol. 60, pp. 1141-1154, 2011.

[6] R. C. Hendriks, R. Heusdens, U. Kjems and J. Jensen, "On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement," *Signal Processing Letters, IEEE,* vol. 16, pp. 885-888, 2009.

[7] Oppenheim, A. V. and Schafer, R. W., *Discrete-Time Signal Processing.* Upper Saddle River, NJ: Pearson Education, Inc., 2010.

[8] P. C. Loizou, *Speech Enhancement: Theory and Practice.* Baca Raton, Fl: CRC Press, 2007.

[9] P. M. Hofman, J. G. A. Van Riswick and A. J. Van Opstal, "Relearning sound localization with new ears," *Relearning Sound Localization with New Ears,* 1998.

[10] M. Raspaud, H. Viste and G. Evangelista, "Binaural Source Localization by Joint Estimation of ILD and ITD," *Audio, Speech, and Language Processing, IEEE Transactions On,* vol. 18, pp. 68-77, 2010.

[11] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann and B. Kollmeier, "Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses ," *EURASIP Journal on Advances in Signal Processing,* vol. 2009, 2009.

[12] V. R. Algazi and R. O. Duda, "Effective use of psychoacoustics in motion-tracked binaural audio," in *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium On,* 2008, pp. 562-567.

[13] D. S. Bernstein, *Matrix Mathematics.* 2009.

[14] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* Academic Press, 1999.

[15] A. Avni and B. Rafaely, "Interaural cross correlation and spatial correlation in a sound field represented by spherical harmonics," *Ambisionics Symposium,* 2009.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* Elsevier Academic Press, 2007.

[17] M. Brandstein and D. Ward, *Microphone Arrays: Singal Processing Techniques and Applications.* Springer, 2001.

[18] E. A. P. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *Journal of the Acoustical Society of America,* vol. 122, pp. 3464-3470, 2007.

[19] M. Jeub, M. Dorbecker and P. Vary, "A Semi-Analytical Model for the Binaural Coherence of Noise Fields," *Signal Processing Letters, IEEE,* vol. 18, pp. 197-200, 2011.

[20] C. Herglotz, M. Jeub, C. Nelke, C. Beaugeant and P. Vary, "Evaluation of single- and dual-channel noise noise power spectral density estimation algorithms for mobile phones," in *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)* Aachen, Germany, 2011, .

[21] R. C. Hendriks, R. Heusdens and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference On,* 2010, pp. 4266-4269.

[22] H. L. Van Trees, *Optimum Array Processing.* John Wiley & Sons, Inc., 2002.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations.* The John Hopkins University Press, 2013.

[24] G. W. Stewart, "On the Sensitivity of the Eigenvalue Problem $Ax = \lambda Bx$," *SIAM Journal on Numerical Analysis,* vol. 9, pp. 669-686, December 1972, 1972.

[25] R. Martin, "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors," *Speech and Audio Processing, IEEE Transactions On,* vol. 13, pp. 845-856, 2005.

[26] T. W. Anderson, "Asymptotic Theory for Principal Component Analysis," *Annals of Mathematical Statistics,* vol. 34, pp. 122-148, Mar., 1963, 1963.

[27] R. P. Gupta, "Asymptotic theory for principal component analysis in the complex case," *J. Indian Stat. Assoc.,* vol. 3, pp. 97-106, 1965.

[28] M. S. Bartlett, "A note on the multiplying factors for various χ^2 approximations," *J. R. Stat. Soc.,* vol. 16, pp. 296-298, 1954.

[29] D. N. Lawley, "Tests of significance of the latent roots of the covariance and correlation matrices," *Biometrica,* vol. 43, pp. 128-136, 1956.

[30] J. A. Uber, "Estimation of the Dimensionality of the Signal Subspace," Thursday, December 4, 2003.

[31] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions On,* vol. 19, pp. 716-723, 1974.

[32] J. Rissanen, "Modeling by shortest data description," *Automatica,* vol. 14, pp. 465-471, 1978.

[33] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.,* vol. 6, pp. 461-464, 1978.

[34] Weisstein, Eric W. "Quartic Equation." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/QuarticEquation.html

[35] L. N. Trefethen and I. David Bau, *Numerical Linear Algebra.* SIAM, 1997.

[36] R. Martin, "Spectral Subtraction based on minimum statistics," *Proc. Eur. Signal Process,* pp. p.1182-1185, 1994.

[37] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *Speech and Audio Processing, IEEE Transactions On,* vol. 9, pp. 504-512, 2001.

[38] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," *Proc. Eurospeech,* pp. 1513-1516, 1995.

[39] R. G. Gallager, *Principles of Digital Communication.* The Edimburgh building: Cambridge University Press, 2008.

[40] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *Audio and Electroacoustics, IEEE Transactions On,* vol. 15, pp. 70-73, 1967.

[41] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions On,* vol. 32, pp. 1109-1121, 1984.

[42] R. C. Hendriks, J. Jensen and R. Heusdens, "Noise Tracking Using DFT Domain Subspace Decompositions," *Audio, Speech, and Language Processing, IEEE Transactions On,* vol. 16, pp. 541-553, 2008.

[43] R. W. D. Nickalls, "Viète, Descartes and the cubic equation," *The Mathematical Gazette,* vol. 90, pp. 203-208, 2006.